Spring 2018

AN INVESTIGATION INTO THE PROPERTIES OF QUATERNIONS: THEIR ORIGIN, BASIC PROPERTIES, FUNCTIONAL ANALYSIS, AND ALGEBRAIC CHARACTERISTICS

James Miller
John Carroll University, jmiller15@jcu.edu

AN INVESTIGATION INTO THE PROPERTIES OF QUATERNIONS: THEIR ORIGIN, BASIC PROPERTIES, FUNCTIONAL ANALYSIS, AND ALGEBRAIC CHARACTERISTICS

An Essay Submitted to the
Office of Graduate Studies
College of Arts & Sciences of
John Carroll University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science

By
James Miller
2018

The essay of James Miller is hereby accepted:

Advisor – Barbara D'Ambrosia   Date

Co-Advisor – Paul Shick   Date

I certify that this is the original document:

Author – James Miller   Date

0. Introduction: Origins of the Quaternion Algebra

In 1833 William Rowan Hamilton presented a paper on complex numbers before the Irish Academy ([1]). For many decades prior, mathematicians throughout Europe were formulating precisely how complex numbers should be understood. Were they geometrical constructs? Were they purely algebraic, the playthings of cloistered mathematicians who preferred the solitude of numbers to the boisterous company of their compatriots? Were they something of both or neither at all?

Hamilton, with keen insight that was typical of the bibliophilic Irish since the days of St. Patrick, seemed to settle the question with an algebraic understanding of complex numbers as multiplication of ordered pairs of real numbers according to the following rule:

$$(x, y)\cdot(w, z) = (xw - yz, xz + yw).$$

Hamilton interpreted this operation as a rotation in the two-dimensional plane (this can be proven by considering the polar representation of the ordered pairs). This interpretation was not entirely new, of course, since the geometrical descriptions of complex numbers provided by mathematicians of a previous generation contained, at least implicitly, this notion of rotation. However, Hamilton's paper made this most explicit and thus continued the marriage of algebra with geometry that Descartes had started nearly two centuries prior.

Encouraged by the reception of his work and recognizing the usefulness of conceiving these mysterious complex numbers in a more concrete way, Hamilton decided to extend his idea of multiplying ordered pairs to multiplying ordered triples. Motivated by the complex numbers, Hamilton defined another imaginary term j so that $i \neq \pm j$ and $j^2 = -1$. He considered an arbitrary triple represented as $p = a + bi + cj$ for real numbers *a*, *b*, and *c* and defined the modulus of *p* analogously to the definition of modulus for complex numbers:

$$|p| = \sqrt{a^2 + b^2 + c^2}\ .$$

Next, he assumed that the modulus of a product is the product of the moduli in order to derive the following for a given triple $p = a + bi + cj$:

$$\begin{aligned} a^2 + b^2 + c^2 &= |p|^2 = |p||p| = |pp| = \left|p^2\right| = \left|(a + bi + cj)^2\right| \\ &= \left|a^2 + abi + acj + bai - b^2 + bcij + caj + cbji - c^2\right| \end{aligned}$$

$$= \left| a^2 - b^2 - c^2 + (ab + ba)i + (ac + ca)\, j + bcij + cbji \right|$$
$$= \left| a^2 - b^2 - c^2 + (2ab)i + (2ac)\, j + bcij + cbji \right|.$$

The vector under his consideration now had a real component, an *i* component, a *j* component, an *ij* component and a *ji* component. Since Hamilton was working in three dimensions and had implicitly defined 1, *i* and *j* to be linearly independent, he had to find a way to incorporate the *ij* and *ji* components into this basis. Hamilton's first attempt was to set $ij = 0$ and to assume that *i* and j commute. Hence he derived

$$\begin{aligned} \left| a^2 - b^2 - c^2 + (2ab)i + (2ac)\, j \right| &= \sqrt{\left(a^2 - b^2 - c^2\right)^2 + (2ab)^2 + (2ac)^2} \\ &= \sqrt{a^4 + b^4 + c^4 + 2a^2b^2 + 2a^2c^2 + 2b^2c^2} \\ &= \sqrt{\left(a^2 + b^2 + c^2\right)^2} = a^2 + b^2 + c^2. \end{aligned}$$

Thus, his assumptions that $ij = 0$ and $ji = 0$ were consistent with his assumption that the modulus of a product is the product of the moduli. However, it did not seem right to Hamilton that $ij = 0$. Therefore, he decided upon something else. Since he still wanted those pesky *ij* and *ji* terms to vanish, he set $ij = -ji$ and pressed onward. This simple mathematical hieroglyph, perhaps so trivial to us now, was quite an earthshattering inscription when Hamilton put his pen to paper. While at the time it was not unheard of to consider that it might be possible to suspend the commutative property of multiplication, no one was quite sure if a coherent and consistent system could be distilled from non-commutative multiplication. Hamilton was setting sail for terra incognita. Perhaps sensing some trepidation at the onset of this journey into the mathematical unknown, he defined a place holder, a mathematical life-preserver, $k = ij$, and then proceeded to derive other properties of his triples, allowing the possibility that *k* may or may not be 0.

Eventually, having made virtually no progress in the three-dimensional world, Hamilton decided to jump into a fourth dimension and allowed *k* to be linearly independent of 1, *i*, and *j*. He proved that $k^2 = -1$ and determined that while the commutative property is lost in his new system, the associativity of multiplication – a word first used by Hamilton himself – remains valid. Hamilton then derived the products of the basis elements 1, *i*, *j*, and *k* that create a consistent system of multiplying 4-tuples, reproduced here as a table for easy visualization:

$\cdot$	i	j	k
i	-1	k	$-j$
j	$-k$	-1	i
k	j	$-i$	-1

Hamilton called this system the *quaternion algebra*, a word he took from the King James Version of Acts 12:4, which, telling of King Herod's arresting of the apostle Peter, says "And when he had apprehended him, he put him in prison, and delivered him to four quaternions of soldiers to keep him; intending after Easter to bring him forth to the people" ([2]). Shortly after he presented his paper on multiplying ordered pairs, Hamilton presented his forays into this new system to the Irish Academy. A man who had already amassed numerous achievements – at the young age of five he had already mastered Latin, Greek, and Hebrew; at the age of 10, he extended his linguistic knowledge to other Eastern languages; and while an undergraduate he had been appointed a professor of astronomy – Hamilton regarded the discovery of quaternions as his greatest accomplishment ([2]). Perhaps rather recklessly he endowed this invention of his with cosmic significance, going so far as to place his development of the quaternions on the same level as Newton's development of Calculus. Other mathematicians with similar illusions of grandeur developed whole societies devoted to the investigation and propagation of the study of quaternions, hoping to discover applications of these mysterious beasts that equaled those of the Calculus. In hindsight, the almost religious fever with which quaternions were hailed seems silly since, of course, their applications – outside a few instances in physics – are very limited. Even in abstract mathematics, a discipline which often seems to delight in oddities, quaternions hardly appear. Nevertheless, Hamilton's discovery in one respect did attain cosmic significance: it paved the way for the development of non-commutative algebras. For nearly all of mathematics prior, the commutativity of multiplication seemed not only obvious but also necessary in order to obtain a consistent system of multiplication. The quaternions demolished this necessity and revealed just how versatile mathematical systems could be.

The present essay is arranged in roughly three sections and seeks to accomplish three things. In the first section, we lay out the basic properties of quaternions and show how to perform basic operations with these numbers. The essay rises in difficulty with its discussion of two theorems: Frobenius's Theorem that the only associative division algebras over the real numbers are the real numbers, the complex numbers, and the quaternions, and a Fundamental Theorem of Algebra for Polynomials with Quaternion Coefficients. The former theorem is well known and serves as a capstone of the essay since it implies that the discussion of this present paper cannot be extended into higher dimensions without sacrificing at least one important property. The latter theorem is almost unheard of and was itself not treated in professional mathematical literature until 1941. Its proof as outlined in [2] is brief. The purpose of our essay is to expand upon the proof immensely. Therefore, we spend quite a bit of time building up the algebraic topological machinery necessary for the proof. The required elements of algebraic topology are found in Hatcher's *Algebraic Topology* ([4]). While we borrow extensively from Hatcher, we provide a much more thorough discussion of many of Hatcher's proofs, which are often quite short, dense, and difficult.

1. Basic Definitions and Operations

We let $\mathbb{H}$ denote the set of quaternions. That is, $\mathbb{H}=\{\alpha+\beta i+\gamma j+\delta k \mid \alpha,\beta,\gamma,\delta\in\mathbb{R}\}$. Addition on $\mathbb{H}$ is defined component-wise so that

$$(\alpha+\beta i+\gamma j+\delta k)+(\varepsilon+\upsilon i+\psi j+\omega k)=(\alpha+\varepsilon)+(\beta+\upsilon)i+(\gamma+\psi)j+(\delta+\omega)k .$$

It is readily apparent that addition over $\mathbb{H}$ inherits commutativity from the real numbers. We define multiplication based on the distributive property and the rules for multiplying the imaginary components of quaternions laid down by Sir Hamilton to guarantee consistency. Without going into the details of the results of distribution, we obtain the following general rule for multiplication of quaternions:

$$\begin{aligned}&(\alpha+\beta i+\gamma j+\delta k)(\varepsilon+\upsilon i+\psi j+\omega k)\\ &\quad=(\alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega)+(\alpha\upsilon+\beta\varepsilon+\gamma\omega-\delta\psi)i\\ &\quad\quad+(\alpha\psi-\beta\omega+\gamma\varepsilon+\delta\upsilon)j+(\alpha\omega+\beta\psi-\gamma\upsilon+\delta\varepsilon)k.\end{aligned}$$

We adopt the following convention for quaternions: we will use Roman letters to denote quaternions and Greek letters to denote the components of a quaternion. So for example, we will write $a=(\alpha,\beta,\gamma,\delta)$ or $a=\alpha+\beta i+\gamma j+\delta k$ where *a*, of course, is a quaternion with components α, β, γ, and δ. When discussing functions of quaternions we will use *q* as a dummy variable; e.g., $f:\mathbb{H}\to\mathbb{H}$ by $f(q)=q^2$.

It is clear that writing out quaternions component-wise is rather bulky and inefficient. We will often write quaternions as ordered quadruples when recording a specific quaternion in a paragraph, so that $\alpha+\beta i+\gamma j+\delta k$ is written $(\alpha,\beta,\gamma,\delta)$. For calculations, however, we will express them variously as row vectors and 4×4 matrices. In this way, we may think of each quaternion as a row vector and define multiplication in terms of matrices. To make this unambiguous, we let $(\alpha,\beta,\gamma,\delta)$ denote a quaternion and let $[\alpha,\beta,\gamma,\delta]$ denote the row vector that corresponds to the previous quaternion when we multiply. Since we know the product of two quaternions must be itself a quaternion, our matrix notion requires that we multiply a row vector by a 4×4 matrix in order to obtain another row vector. That is, for quaternions $(\alpha,\beta,\gamma,\delta),(\varepsilon,\upsilon,\psi,\omega)\in\mathbb{H}$,

$$(\alpha,\beta,\gamma,\delta)\cdot(\varepsilon,\upsilon,\psi,\omega)=[\alpha,\beta,\gamma,\delta]\begin{bmatrix}\varepsilon & \upsilon & \psi & \omega\\ -\upsilon & \varepsilon & -\omega & \psi\\ -\psi & \omega & \varepsilon & -\upsilon\\ -\omega & -\psi & \upsilon & \varepsilon\end{bmatrix}=\begin{bmatrix}\alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega\\ \alpha\upsilon+\beta\varepsilon+\gamma\omega-\delta\psi\\ \alpha\psi-\beta\omega+\gamma\varepsilon+\delta\upsilon\\ \alpha\omega+\beta\psi-\gamma\upsilon+\delta\varepsilon\end{bmatrix}^T .$$

Example 1: Let us multiply the quaternions $(1,2,3,4)$ and $(4,3,2,1)$ using the matrix notation:

$$(1,2,3,4)(4,3,2,1)=[1,2,3,4]\begin{bmatrix} 4 & 3 & 2 & 1 \\ -3 & 4 & -1 & 2 \\ -2 & 1 & 4 & -3 \\ -1 & -2 & 3 & 4 \end{bmatrix}$$

$$=\begin{bmatrix} 4-6-6-4 \\ 3+8+3-8 \\ 2-2+12+12 \\ 1+4-9+16 \end{bmatrix}^T=\begin{bmatrix} -12 \\ 6 \\ 24 \\ 12 \end{bmatrix}^T=(-12,6,24,12).$$

To demonstrate the non-commutativity of multiplication, we reverse the order:

$$(4,3,2,1)(1,2,3,4)=[4,3,2,1]\begin{bmatrix} 1 & 2 & 3 & 4 \\ -2 & 1 & -4 & 3 \\ -3 & 4 & 1 & -2 \\ -4 & -3 & 2 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} 4-6-6-4 \\ 8+3+8-3 \\ 12-12+2+2 \\ 16+9-4+1 \end{bmatrix}^T=\begin{bmatrix} -12 \\ 16 \\ 4 \\ 22 \end{bmatrix}^T=(-12,16,4,22).$$

Next we prove a very useful theorem for this essay. This theorem translates the properties of quaternions to the properties of matrices with which we are presumably better acquainted.

Theorem 1: Let $\mathrm{H}=\left\{\begin{bmatrix} w & -z \\ \overline{z} & \overline{w} \end{bmatrix} \middle| w,z\in\mathbb{C}\right\}$. There exists a ring isomorphism $\Phi:\mathbb{H}\to\mathrm{H}$ defined by $\Phi((\alpha,\beta,\gamma,\delta))=\begin{bmatrix} \alpha+\beta i & -\gamma-\delta i \\ \gamma-\delta i & \alpha-\beta i \end{bmatrix}$.

Proof: It is straightforward to see that Φ is a function from $\mathbb{H}$ to $\mathrm{M}_{2\times 2}(\mathbb{C})$ and that $\mathrm{H}=\Phi(\mathbb{H})$. So, provided that Φ is a homomorphism, H is a subring of $\mathrm{M}_{2\times 2}(\mathbb{C})$, and is thus a ring.

Now, let $a,b\in\mathbb{H}$ with $a=(\alpha,\beta,\gamma,\delta)$ and $b=(\varepsilon,\upsilon,\psi,\omega)$. Then,

$$\Phi(a)=\begin{bmatrix}\alpha+\beta i & -\gamma-\delta i\\ \gamma-\delta i & \alpha-\beta i\end{bmatrix} \text{ and } \Phi(b)=\begin{bmatrix}\varepsilon+\upsilon i & -\psi-\omega i\\ \psi-\omega i & \varepsilon-\upsilon i\end{bmatrix}.$$

And so,

$$\begin{aligned}\Phi(a)\Phi(b)\\ &=\begin{bmatrix}(\alpha+\beta i)(\varepsilon+\upsilon i)+(-\gamma-\delta i)(\psi-\omega i) & (\alpha+\beta i)(-\psi-\omega i)+(-\gamma-\delta i)(\varepsilon-\upsilon i)\\ (\gamma-\delta i)(\varepsilon+\upsilon i)+(\alpha-\beta i)(\psi-\omega i) & (\gamma-\delta i)(-\psi-\omega i)+(\alpha-\beta i)(\varepsilon-\upsilon i)\end{bmatrix}\\ &=\begin{bmatrix}w & -z\\ \bar{z} & \bar{w}\end{bmatrix},\end{aligned}$$

where

$$w=(\alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega)+(\alpha\upsilon+\beta\varepsilon+\gamma\omega-\delta\psi)i$$

and

$$z=(\alpha\psi-\beta\omega+\gamma\varepsilon+\delta\upsilon)+(\alpha\omega+\beta\psi-\gamma\upsilon+\delta\varepsilon)i.$$

By the definition of multiplication in $\mathbb{H}$,

$$ab=(\alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega,\ \alpha\upsilon+\beta\varepsilon+\gamma\omega-\delta\psi,\ \alpha\psi-\beta\omega+\gamma\varepsilon+\delta\upsilon,\ \alpha\omega+\beta\psi-\gamma\upsilon+\delta\varepsilon).$$

And so,

$$\Phi(ab)=\begin{bmatrix}x & -y\\ \bar{y} & \bar{x}\end{bmatrix}$$

where

$$x=(\alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega)+(\alpha\upsilon+\beta\varepsilon+\gamma\omega-\delta\psi)i$$

and

$$y=(\alpha\psi-\beta\omega+\gamma\varepsilon+\delta\upsilon)+(\alpha\omega+\beta\psi-\gamma\upsilon+\delta\varepsilon)i.$$

Comparing these two results, we see that $\Phi(ab)=\Phi(a)\Phi(b)$. It is rather trivial to see that $\Phi(a+b)=\Phi(a)+\Phi(b)$ since this follows from matrix addition. And so, $\Phi:\mathbb{H}\to\mathrm{H}$ is a surjective ring homomorphism. Finally, it is easy to see that Φ is also injective and thus is a ring isomorphism. ■

A simple corollary of Theorem 1 is that multiplication in $\mathbb{H}$ is not commutative (as we already saw with Example 1). A second corollary of Theorem 1 is that each non-zero quaternion has an inverse. We may use the function Φ to derive the general form of an inverse. Let $a \in \mathbb{H} - \{0\}$, with $a = (\alpha, \beta, \gamma, \delta)$. Then, since Φ is a ring isomorphism

$$\Phi\left(a^{-1}\right) = \Phi(a)^{-1} = \begin{bmatrix} \alpha + \beta i & -\gamma - \delta i \\ \gamma - \delta i & \alpha - \beta i \end{bmatrix}^{-1} = \left(\frac{1}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2} \begin{bmatrix} \alpha - \beta i & \gamma + \delta i \\ -\gamma + \delta i & \alpha + \beta i \end{bmatrix} \right).$$

Therefore,

$$a^{-1} = \left(\frac{\alpha}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}, \frac{-\beta}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}, \frac{-\gamma}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}, \frac{-\delta}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2} \right).$$

We now proceed to derive some properties of quaternions that will be of use later and are interesting in their own right. We begin with some basic definitions.

Definition 1: We define the *imaginary space* of quaternions to be

$$\operatorname{Im}\mathbb{H} = \{(\alpha, \beta, \gamma, \delta) \in \mathbb{H} \mid \alpha = 0\}.$$

The imaginary space is aptly named, for it has similar properties to the imaginary space in $\mathbb{C}$, the complex plane. For $a \in \operatorname{Im}\mathbb{H}$, $a = \beta i + \gamma j + \delta k$, and

$$a^2 = [0, \beta, \gamma, \delta] \begin{bmatrix} 0 & \beta & \gamma & \delta \\ -\beta & 0 & -\delta & \gamma \\ -\gamma & \delta & 0 & -\beta \\ -\delta & -\gamma & \beta & 0 \end{bmatrix} = \begin{bmatrix} -\beta^2 - \gamma^2 - \delta^2 \\ \gamma\delta - \delta\gamma \\ -\beta\delta + \delta\beta \\ \beta\gamma - \gamma\beta \end{bmatrix}^T = \begin{bmatrix} -\beta^2 - \gamma^2 - \delta^2 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T.$$

Hence, a^2 is a non-positive real number.

For comparison the square of a general quaternion $a = (\alpha, \beta, \gamma, \delta)$ is

$$a^2 = \left(\alpha^2 - \beta^2 - \gamma^2 - \delta^2, 2\alpha\beta, 2\alpha\gamma, 2\alpha\delta\right).$$

Definition 2: We define the *conjugate* of a quaternion $a = (\alpha, \beta, \gamma, \delta)$ to be

$$\bar{a} = (\alpha, -\beta, -\gamma, -\delta).$$

We see that the conjugate of a quaternion has a property analogous to its complex counterpart:

$$\begin{aligned} a\bar{a} &= \left(\alpha^2+\beta^2+\gamma^2+\delta^2\right)+\left(-\alpha\beta+\beta\alpha-\gamma\delta+\delta\gamma\right)i \\ &\quad +\left(-\alpha\gamma+\beta\delta+\gamma\alpha-\delta\beta\right)j+\left(-\alpha\delta-\beta\gamma+\gamma\beta+\delta\alpha\right)k \\ &= \alpha^2+\beta^2+\gamma^2+\delta^2. \end{aligned}$$

We also note that $\bar{a}a$ yields the same product. However, instead of checking this product by multiplying, we conclude this fact in a corollary to following theorem.

Theorem 2: The center of $\mathbb{H}$, denoted by $Z\left(\mathbb{H}\right)$, is $\mathbb{R}$.

Proof: Let $x=\left(\alpha,\beta,\gamma,\delta\right)\in\mathbb{H}$ and suppose $x\in Z\left(\mathbb{H}\right)$. Then, for all $u=\left(\varepsilon,\upsilon,\psi,\omega\right)\in\mathbb{H}$, $xu=ux$. So

$$\begin{bmatrix} \alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega \\ \alpha\upsilon+\beta\varepsilon-\delta\psi+\gamma\omega \\ \alpha\psi-\beta\omega+\delta\upsilon+\gamma\varepsilon \\ \alpha\omega+\beta\psi+\delta\varepsilon-\gamma\upsilon \end{bmatrix}^T = \begin{bmatrix} \alpha\varepsilon-\beta\upsilon-\gamma\psi-\delta\omega \\ \alpha\upsilon+\beta\varepsilon-\gamma\omega+\delta\psi \\ \alpha\psi+\beta\omega-\delta\upsilon+\gamma\varepsilon \\ \alpha\omega-\beta\psi+\delta\varepsilon+\gamma\upsilon \end{bmatrix}^T.$$

Equating the second, third, and fourth entries of the two row vectors, we have the following three equations:

$$-\left(\delta\psi-\gamma\omega\right)=\left(\delta\psi-\gamma\omega\right),\ -\left(\beta\omega-\delta\upsilon\right)=\left(\beta\omega-\delta\upsilon\right),\ -\left(\beta\psi-\gamma\upsilon\right)=\beta\psi-\gamma\upsilon.$$

Thus, $\delta\psi=\gamma\omega,\ \beta\omega=\delta\upsilon,$ and $\beta\psi=\gamma\upsilon$. In particular, $\dfrac{\beta}{\upsilon}=\dfrac{\gamma}{\psi}=\dfrac{\delta}{\omega}$ whenever υ, ψ, and ω are nonzero. Since υ, ψ, and ω are arbitrary real numbers (and hence nonzero in some instances), it follows that $\beta=0$, $\gamma=0$, and $\delta=0$.

Thus $x\in\mathbb{R}$, and so $Z\left(\mathbb{H}\right)\subseteq\mathbb{R}$. The reverse inclusion is rather trivial and so we conclude $Z\left(\mathbb{H}\right)=\mathbb{R}$. ■

Corollary 1: We see that for a fixed $a=\left(\alpha,\beta,\gamma,\delta\right)\in\mathbb{H}$, the centralizer of *a* is

$$Z\left(\{a\}\right)=\left\{\left(\varepsilon,\upsilon,\psi,\omega\right)\in\mathbb{H}\,\middle|\,\delta\psi=\gamma\omega,\beta\omega=\delta\upsilon,\text{ and }\beta\psi=\gamma\upsilon\right\}.$$

We have not reproduced the ratios from the proof of Theorem 2 in this definition of centralizer so as to avoid cases where one of the denominators is 0. However, the ratios are quite useful when they are defined.

In particular, $\bar{a} \in Z(\{a\})$ since $\frac{\beta}{\upsilon} = \frac{\gamma}{\psi} = \frac{\delta}{\omega} = -1$ in this case. Hence, we complete the proof from above that $a\bar{a} = \bar{a}a = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$. ■

Example 2: Note that $(5,12,16,2)$ and $(17,24,32,4)$ commute by the converse of the ratio argument in Theorem 2 since the common ratio among the second, third, and fourth components is $\frac{1}{2}$.

Definition 3: Let $\mathbb{C}_i = \{(\alpha,\beta,\gamma,\delta) \in \mathbb{H} \mid \gamma = \delta = 0\}$, $\mathbb{C}_j = \{(\alpha,\beta,\gamma,\delta) \in \mathbb{H} \mid \beta = \delta = 0\}$ and $\mathbb{C}_k = \{(\alpha,\beta,\gamma,\delta) \in \mathbb{H} \mid \beta = \gamma = 0\}$.

Theorem 3: The sets $\mathbb{C}_i$, $\mathbb{C}_j$, and $\mathbb{C}_k$ are subrings of $\mathbb{H}$ and further $\mathbb{C}_i \cong \mathbb{C}_j \cong \mathbb{C}_k$.

Proof: A routine check shows that the sets $\mathbb{C}_i$, $\mathbb{C}_j$, and $\mathbb{C}_k$ are subrings of $\mathbb{H}$. Now let $f : \mathbb{C}_i \to \mathbb{C}_j$ by $f(\alpha + \beta i) = \alpha + \beta j$. It is easy to see that *f* is bijective. We see that *f* is a homomorphism since

$$\begin{aligned} f\left((\alpha+\beta i)(\gamma+\delta i)\right) &= f\left((\alpha\gamma-\beta\delta)+(\alpha\delta+\beta\gamma)i\right) \\ &= (\alpha\gamma-\beta\delta)+(\alpha\delta+\beta\gamma)j \\ &= (\alpha+\beta j)(\gamma+\delta j) \\ &= f(\alpha+\beta i)\, f(\gamma+\delta i). \end{aligned}$$

The same results can easily be established for $g : \mathbb{C}_j \to \mathbb{C}_k$ defined by

$$g(\alpha + \gamma j) = \alpha + \gamma k$$

and $h : \mathbb{C}_i \to \mathbb{C}_k$ defined by $h(\alpha + \beta i) = \alpha + \beta k$. ■

We have already established that $a\bar{a} \in \mathbb{R}$. It is only natural to consider whether quaternion conjugates have other properties analogous to those of complex conjugates.

Theorem 4:

(1) For all $a, b \in \mathbb{H}$ $\overline{a+b} = \bar{a} + \bar{b}$.
(2) For some $a, b \in \mathbb{H}$ $\overline{ab} \neq \bar{a}\bar{b}$.

Proof: Let $a = (\alpha,\beta,\gamma,\delta)$ and $b = (\varepsilon,\upsilon,\psi,\omega)$. Then,

$$\begin{aligned}\overline{a}+\overline{b} &= (\alpha,-\beta,-\gamma,-\delta)+(\varepsilon,-\upsilon,-\psi,-\omega)=(\alpha+\varepsilon,-\beta-\upsilon,-\gamma-\psi,-\delta-\omega)\\ &= \overline{(\alpha+\varepsilon,\beta+\upsilon,\gamma+\psi,\delta+\omega)}=\overline{a+b}.\end{aligned}$$

So (1) holds. For (2) we consider an example. Let $a=(1,2,3,4)$ and $b=(4,3,2,1)$. From Example 1, we see that $\overline{ab}=\overline{(-12,6,24,12)}=(-12,-6,-24,-12)$. Whereas,

$$\overline{a}\overline{b}=(1,-2,-3,-4)(4,-3,-2,-1)=\begin{bmatrix}1 & -2 & -3 & -4\end{bmatrix}\begin{bmatrix}4 & -3 & -2 & -1\\ 3 & 4 & 1 & -2\\ 2 & -1 & 4 & 3\\ 1 & 2 & -3 & 4\end{bmatrix}$$

$$=\begin{bmatrix}4-6-6-4\\ -3-8+3-8\\ -2-2-12+12\\ -1+4-9-16\end{bmatrix}^T=\begin{bmatrix}-12\\ -16\\ -4\\ -22\end{bmatrix}^T.$$

Therefore, (2) holds. ■

Example 3: Let $a\in\mathbb{R}$ and $b=(\varepsilon,\upsilon,\psi,\omega)\in\mathbb{H}$. Then $\overline{ab}=\overline{a}\overline{b}=a\overline{b}$ and $\overline{ba}=\overline{b}\overline{a}=\overline{b}a$.

One of the implications of statement (2) in Theorem 4 is that we cannot extend Theorem 1 to higher dimensions. That is, we cannot define a homomorphism $\Psi:U\to O$ where $U=\left\{\begin{bmatrix}a & -b\\ \overline{b} & \overline{a}\end{bmatrix}\middle| a,b\in\mathbb{H}\right\}$ and O is some set of 8-tuples. Specifically, U is not a ring, since, for quaternions a, b, c, and d, the product

$$\begin{bmatrix}a & -b\\ \overline{b} & \overline{a}\end{bmatrix}\begin{bmatrix}c & -d\\ \overline{d} & \overline{c}\end{bmatrix}=\begin{bmatrix}ac-b\overline{d} & -ad-b\overline{c}\\ \overline{b}c+\overline{a}\overline{d} & -\overline{b}d+\overline{a}\,\overline{c}\end{bmatrix}$$

may not be an element of U because the conjugate of the upper left entry of the product may not equal the lower right entry of the product. We will see when we prove Frobenius's Theorem that we have no hope of defining an isomorphism from matrices with quaternion entries to any set of 8-tuples.

2. Fundamental Theorem of Algebra for Quaternions

Although we ended the previous section with a reference to Frobenius's Theorem, we postpone the proof of that famous theorem. Instead we will look at something analogous to the Fundamental Theorem of Algebra for polynomials over the quaternions. First, let us consider some examples of functions defined over the quaternions.

Example 4: Let $f(q) = q^2 + 1$. Then, some roots of $f(q)$ in $\mathbb{H}$ are $\pm i$, $\pm j$, and $\pm k$. Thus, $f(q)$ has at least six roots. However, recall from our discussion of the imaginary space of $\mathbb{H}$ that an element $a \in \mathrm{Im}\mathbb{H}$ satisfies $a^2 \in \mathbb{R}$ and $a^2 \le 0$. Hence, any $a \in \mathrm{Im}\mathbb{H}$ such that $a^2 = -1$ is also a root of *f*. And so, f has uncountably many roots over $\mathbb{H}$.

Example 5: Let $g(q) = iq - qi$. Then, clearly $g(q) = 0$ identically over $\mathbb{C}_i$, but can be non-zero over $\mathbb{H}$. For example,

$$g(j) = ij - ji = 2k \neq 0.$$

From these two examples alone it is enough to see why a Fundamental Theorem of Algebra for polynomials over the quaternions is going to be a statement very different from its counterpart in complex analysis or abstract algebra.

And so we now come to a unique and somewhat obscure result in this essay. As we said in the introduction, this is the most abstract and technical part of the essay. We refer to Hatcher's *Algebraic Topology* ([4]) for the following discussion. We borrow the notation from Hatcher and mine the book for as much as we can. We provide necessary elaboration, given Hatcher's brief and dense style.

To begin, recall that an *n*-dimensional simplex is the smallest convex subset of *n*-dimensional Euclidean space that contains $n+1$ *linearly independent* points. A 0-simplex is a point, a 1-simplex is a line segment (containing the two endpoints), a 2-simplex is a triangle, and a 3-simplex is a tetrahedron. In other words, an *n*-simplex is an *n*-dimensional tetrahedron.

Note that an *n*-simplex is determined by its $n+1$ vertices. Let $v_0, v_1, \ldots, v_n$ be these vertices and denote the simplex by $\Delta^n = [v_0, v_1, \ldots, v_n]$.

Let *X* be a space. We define embeddings of simplicies into *X* as functions $\sigma_\alpha^n : \Delta^n \to X$ where α is an element of an arbitrary indexing set Λ and the superscript *n* indicates that the embedding maps an *n*-simplex into *X*. In particular, we define the embeddings so that $X = \bigcup_{\alpha \in \Lambda,\, i \in \mathbb{N}} \sigma_\alpha^i$. In this way σ_α^n "bends" and "stretches" (in a continuous manner, of course) an n-simplex in order to build the space X. From a somewhat crude perspective, we see the embeddings σ_α^n provide a means to "count" the number of n-simplices required to construct the space X. Further, the bending and stretching of these embeddings provide information on the structure of X. The concept of homology tries to make these notions of the number of embeddings and the structure of X more precise.

Denote by $C_n(X)$ the free abelian group for which each embedding σ_α^n is a generator. That is,

$$C_n(X) = \left\{ \sum_{i=1}^{m} p_i \sigma_\alpha^n \middle| m \in \mathbb{N}, p_i \in \mathbb{Z}, \text{ and } \sigma_\alpha^n : \Delta^n \to X \text{ is an embedding} \right\}.$$

In other words, $C_n(X)$ is the set of linear combinations of generators for the space X. We point out explicitly that the operation on $C_n(X)$ is simply addition of linear combinations of generators.

Let X and Y be spaces and $f : X \to Y$ be a continuous map. Then for each n, f induces maps $f_\#^n : C_n(X) \to C_n(Y)$. The map $f_\#$ is defined on generators $\sigma_\alpha^n \in C_n(X)$ by $f_\#(\sigma_\alpha^n) = f(\sigma_\alpha^n)$. We also define $f_\#$ to extend linearly over sums of the generators. Let $\omega = \sum_{i=1}^{m} p_i \sigma_\alpha^n \in C_n(X)$. Then

$$f_\#^n(\omega) = f_\#^n\left(\sum_{i=1}^{m} p_i \sigma_\alpha^n \right) \equiv \sum_{i=1}^{m} p_i f_\#^n(\sigma_\alpha^n) = \sum_{i=1}^{m} p_i f(\sigma_\alpha^n).$$

In other words, $f_\#^n$ sends generators of $C_n(X)$ to generators of $C_n(Y)$. This definition ensures that $f_\#^n$ is a homomorphism. Since this result is important, we consider it as a theorem in itself.

Theorem 5: $f_\#^n$ is a homomorphism.

As a final note, we will often omit the superscript n for the generator σ_α^n when it is clear that σ_α^n is a generator of $C_n(X)$. Similarly we will often omit the index α and simply write σ to indicate an arbitrary generator. We included both scripts above for the sake of thoroughness and clarity. We now introduce the boundary map.

Definition 4: Let $\partial_n^X : C_n(X) \to C_{n-1}(X)$ be defined as follows

$$\partial_n^X\left(\sigma_\alpha\right)=\sum_{i=0}^{n}(-1)^n\,\sigma_\alpha\big|_{[v_0,v_1,\ldots,\hat{v}_i,\ldots,v_n]}.$$

We will often omit the *X* superscript when the domain of the boundary map is clear.

Now, what precisely does this mean? The boundary map is defined as a homomorphism on the embeddings of *n*-simplices. The boundary map sends embeddings of *n*-simplices to a sum of embeddings of $(n-1)$-simplices. That is, the boundary map takes us down by one dimension. We begin with an *n*-dimensional object and "strip away" its interior in order to obtain an $(n-1)$-dimensional object. For example, if we start with a tetrahedron, the boundary map takes this tetrahedron and maps it to the "sum" of the four triangles that make up the two-dimensional faces of the tetrahedron. If we have a triangle, the boundary map maps this triangle to the "sum" of the three line segments that make up the boundary of the triangle. But how is this mapping achieved? First, we introduce the "hat" notation. The "hat" notation signifies "remove this vertex from the simplex." That is, if $[v_0,v_1,\ldots,v_n]$ is an *n*-simplex, then $[v_0,v_1,\ldots,\hat{v}_i,\ldots v_n]$ is the $(n-1)$-simplex obtained from the original *n*-simplex by removing the *i*th vertex. Then, we restrict the map σ_α to this newly obtained $(n-1)$-simplex. Finally, we alternate the signs in the sum in order to take into consideration the orientation of the boundary.

We expand upon Hatcher's proof of the very important fact that $\partial_n\circ\partial_{n+1}=0$.

Theorem 6: The composition map $\partial_n\circ\partial_{n+1}:C_{n+1}(X)\xrightarrow{\partial_{n+1}}C_n(X)\xrightarrow{\partial_n}C_{n-1}(X)$ is identically 0.

Proof: Let $\sigma\in C_{n+1}(X)$. Then

$$\partial_{n+1}(\sigma)=\sum_{i=0}^{n+1}(-1)^i\,\sigma\big|_{[v_0,\ldots,\hat{v}_i,\ldots,v_{n+1}]}.$$

Now, for each *i*, $\sigma\big|_{[v_0,\ldots,\hat{v}_i,\ldots,v_{n+1}]}\in C_n(X)$. Therefore, we may apply the *n*-dimensional boundary homomorphism to obtain

$$\partial_n\left(\partial_{n+1}(\sigma)\right)=\partial_n\left(\sum_{i=0}^{n+1}(-1)^i\,\sigma\big|_{[v_0,\ldots,\hat{v}_i,\ldots,v_{n+1}]}\right)=\sum_{i=0}^{n+1}(-1)^i\,\partial_n\left(\sigma\big|_{[v_0,\ldots,\hat{v}_i,\ldots,v_{n+1}]}\right).$$

The slightly tricky part of this proof is figuring out how to remove the second vertex. Suppose, for instance, that the vertex v_i has already been removed. Then, obviously, we cannot remove it again, and so we have to remove another vertex v_j. The index *j* can be either less than or greater than the index *i*. Therefore, we express the above sum as

$$\sum_{i=0}^{n+1}(-1)^i \partial_n\left(\sigma|_{[v_0,\dots,\hat{v}_i,\dots,v_{n+1}]}\right)$$
$$=\underbrace{\sum_i(-1)^i\sum_{j<i}(-1)^j\sigma|_{[v_0,\dots,\hat{v}_j,\dots,\hat{v}_i,\dots,v_{n+1}]}}_{sumA}+\underbrace{\sum_i(-1)^i\sum_{j>i}(-1)^{j-1}\sigma|_{[v_0,\dots,\hat{v}_i,\dots,\hat{v}_j,\dots,v_{n+1}]}}_{sumB}$$

where, for the sake of easy reference, we refer to the first summand as *sumA* and the second summand as *sumB.*

Notice that we have erased the upper and lower limits of the index *i* that were present in the first summation. This is to emphasize that some *i*'s are inapplicable due to the restrictions on the *j*'s. For example, in *sumA* we do not let $i=0$ because there can be no *j*'s less than 0.

Notice further that in *sumB* the exponent for the second -1 is $j-1$. Why do we subtract 1? In short, because we have to take into consideration that we are skipping over the already removed v_i vertex. In the case where $j<i$, we may apply the boundary homomorphism simply by definition. For a fixed *i*, we apply the boundary map only considering those indices $j<i$:

$$\partial\left(\sigma|_{[v_0,v_1,\dots,v_{i-1},\dots]}\right)=\sum_{j<i}(-1)^j\sigma|_{[v_0,\dots,\hat{v}_j,\dots,v_{i-1},\dots]}.$$

Now, if we did not already remove v_i, we would consider the case $j=i$. However, since we have already removed v_i, we must skip that case and go from $j=i-1$ to $j=i+1$. The exponent of -1 does not know that we have made this jump, so it would be wrong to leave the exponent as *j*. Rather, the exponent must be adjusted by subtracting 1. In other words, once we skip v_i, we remove the v_j vertex at the $(j-1)$-th step. Hence, we see why the exponent is $j-1$.

Suppose the summation has reached some fixed *i*. Then, choose some $j<i$. These *i* and *j* indices determine a term $(-1)^i(-1)^j\sigma|_{[v_0,\dots,\hat{v}_j,\dots,\hat{v}_i,\dots,v_{n+1}]}$ in *sumA*. There must be a corresponding term in *sumB* with the same indices *i* and *j* in reversed roles. That is, the *i* term in *sumA* will become the *j* term in *sumB* and the *j* term in *sumA* will become the *i* term in *sumB*. Hence, we may rewrite the summation of *sumA* and *sumB* by grouping together the corresponding terms in *sumA* and *sumB*. For the sake of thoroughness and

clarity, we write (i_a, j_a) to express the indices from *sumA* and (j_b, i_b) to express the indices in *sumB* that correspond as described to the indices in *sumA*. So, since $i_a = j_b$ and $j_a = i_b$, we derive:

$$\begin{aligned}
&\underbrace{\sum_i (-1)^i \sum_{j<i} (-1)^j \sigma|_{[v_0,\ldots,\hat{v}_j,\ldots,\hat{v}_i,\ldots,v_{n+1}]}}_{sumA} + \underbrace{\sum_i (-1)^i \sum_{j>i} (-1)^{j-1} \sigma|_{[v_0,\ldots,\hat{v}_i,\ldots,\hat{v}_j,\ldots,v_{n+1}]}}_{sumB} \\
&\quad = \sum_{(i_a,j_a)=(j_b,i_b)} \left((-1)^{i_a} (-1)^{j_a} \sigma|_{[v_0,\ldots,\hat{v}_{j_a},\ldots,\hat{v}_{i_a},\ldots,v_{n+1}]} + (-1)^{j_b} (-1)^{i_b-1} \sigma|_{[v_0,\ldots,\hat{v}_{j_b},\ldots,\hat{v}_{i_b},\ldots,v_{n+1}]} \right) \\
&\quad = \sum_{(i_a,j_a)=(j_b,i_b)} 0 = 0.
\end{aligned}$$

Thus, we achieve the desired result that the composition is 0. ■

Since $\partial_n \circ \partial_{n+1} = 0$, we see that $\operatorname{im} \partial_{n+1} \subseteq \ker \partial_n$. Therefore, the quotient $\ker \partial_n / \operatorname{im} \partial_{n+1}$ is defined. We call this quotient the *n*-dimensional homology group for the space *X*.

Definition 5: The *n*-dimensional homology group of a space X is

$$H_n(X) = \ker \partial_n / \operatorname{im} \partial_{n+1} .$$

We build up some more machinery before we actually compute any homology groups.

Theorem 7: Let $f : X \to Y$ be continuous. The induced maps $f_\#^n$ and $f_\#^{n-1}$ and the boundary maps ∂_n^X and ∂_n^Y satisfy the relation $f_\#^{n-1} \circ \partial_n^X = \partial_n^Y \circ f_\#^n$. In other words, the diagram below commutes:

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C_{n+1}(X) & \xrightarrow{\partial_{n+1}^X} & C_n(X) & \xrightarrow{\partial_n^X} & C_{n-1}(X) & \xrightarrow{\partial_{n-1}^X} & \cdots \\
 & & \downarrow f_\#^{n+1} & & \downarrow f_\#^n & & \downarrow f_\#^{n-1} & & \\
\cdots & \longrightarrow & C_{n+1}(Y) & \xrightarrow{\partial_{n+1}^Y} & C_n(Y) & \xrightarrow{\partial_n^Y} & C_{n-1}(Y) & \xrightarrow{\partial_{n-1}^Y} & \cdots
\end{array}$$

Proof: The proof is essentially an application of the definitions of the various mappings. For a generator $\sigma_\alpha \in C_n(X)$,

$$\begin{aligned}
\left(f_\#^{n-1} \circ \partial_n^X \right)(\sigma_\alpha) &= f_\#^{n-1}\left(\partial_n^X (\sigma_\alpha) \right) \\
&= f_\#^{n-1}\left(\sum_{i=0}^{n} (-1)^i \sigma_\alpha|_{[v_0,v_1,\ldots,\hat{v}_i,\ldots,v_n]} \right) \\
&= \sum_{i=0}^{n} (-1)^i f_\#^{n-1}\left(\sigma_\alpha|_{[v_0,v_1,\ldots,\hat{v}_i,\ldots,v_n]} \right).
\end{aligned}$$

The key to this proof is the argument that *first restricting* σ_α *to the* $(n-1)$*-simplex* $[v_0,\dots,\hat{v}_i,\dots,v_n]$ *and then applying* $f_\#^{n-1}$ *is the same as first applying* $f_\#^n$ *to* σ_α *and then restricting that composition to the* $(n-1)$*-simplex.* This is easy since

$$\begin{aligned} f_\#^{n-1}\left(\sigma_\alpha\big|_{[v_0,\dots,\hat{v}_i,\dots,v_n]}\right) &= f\left(\sigma_\alpha\big|_{[v_0,\dots,\hat{v}_i,\dots,v_n]}\right) \\ &= f\left(\sigma_\alpha\right)\big|_{[v_0,\dots,\hat{v}_i,\dots,v_n]} \\ &= f_\#^n\left(\sigma_\alpha\right)\big|_{[v_0,\dots,\hat{v}_i,\dots,v_n]} \end{aligned}$$

where the second equality follows from the continuity of *f*.

If we first restrict σ_α, we allow f to map only the $(n-1)$-simplex into the space *Y*, thereby obtaining an $(n-1)$-simplex in *Y*. If, however, we first map the entire *n*-simplex into *Y* and then place this image under the same restriction we obtain an $(n-1)$-simplex in *Y*. Since the two resultant $(n-1)$-simplices are obtained from the same *n*-simplex in *X*, they must be the same.

Hence, the remainder of the proof follows at once:

$$\begin{aligned} \sum_{i=0}^{n}(-1)^i f_\#^{n-1}\left(\sigma_\alpha\big|_{[v_0,v_1,\dots,\hat{v}_i,\dots,v_n]}\right) &= \sum_{i=0}^{n}(-1)^i f_\#^n\left(\sigma_\alpha\right)\big|_{[v_0,v_1,\dots,\hat{v}_i,\dots,v_n]} \\ &= \partial_n^Y\left(f_\#^n\left(\sigma_\alpha\right)\right) = \left(\partial_n^Y \circ f_\#^n\right)\left(\sigma_\alpha\right). \quad \blacksquare \end{aligned}$$

Definition 6: Define $f_*^n : H_n(X) \to H_n(Y)$ by $f_*^n(\Omega) = f_\#^n(\omega)\mathrm{im}\left(\partial_{n+1}^Y\right)$ where $\Omega \in H_n(X)$ such that $\Omega = \omega\mathrm{im}\left(\partial_{n+1}^X\right)$ for some $\omega \in \ker\partial_n^X$. This map is the *induced homomorphism on homology groups* as we will see below.

Note $\partial_n^X(\omega) = 0$ since $\omega \in \ker\partial_n^X$. And so, $0 = f_\#^{n-1}(0) = f_\#^{n-1}\left(\partial_n^X(\omega)\right) = \partial_n^Y\left(f_\#^n(\omega)\right)$. Therefore, $f_\#^n(\omega) \in \ker\partial_n^Y$. And we can conclude that

$$f_*^n(\Omega) \in \ker\left(\partial_n^Y\right)\big/\mathrm{im}\left(\partial_{n+1}^Y\right) = H_n(Y).$$

The map, then, is defined with appropriate domain and codomain.

Theorem 8: The function f_*^n is a homomorphism between homology groups.

Proof: This result follows from the fact that since $f_\#^n$ is a homomorphism,

$$f_*^n(\Omega)+f_*^n(\Psi)=f_\#^n(\omega)\operatorname{im}\left(\partial_{n+1}^Y\right)+f_\#^n(\psi)\operatorname{im}\left(\partial_{n+1}^Y\right)$$
$$=\left(f_\#^n(\omega)+f_\#^n(\psi)\right)\operatorname{im}\left(\partial_{n+1}^Y\right)=f_\#^n(\omega+\psi)\operatorname{im}\left(\partial_{n+1}^Y\right)=f_*^n(\Omega+\Psi). \quad \blacksquare$$

Theorem 9: Let $f:X\to Y$ and $g:Y\to Z$. Then $(g\circ f)_*=g_*\circ f_*$.

Proof: First we note the following mapping diagrams:

$$C_n(X)\xrightarrow{f_\#}C_n(Y)\xrightarrow{g_\#}C_n(Z) \text{ and } H_n(X)\xrightarrow{f_*}H_n(Y)\xrightarrow{g_*}H_n(Z).$$

We observe from the definition of $f_\#$ and $g_\#$ that for $\psi=\sum_i n_i\sigma_i\in C_n(X)$,

$$(g_\#\circ f_\#)(\psi)=g_\#\left(f_\#(\psi)\right)=g_\#\left(f_\#\left(\sum_i n_i\sigma_i\right)\right)=g_\#\left(\sum_i n_i f_\#(\sigma_i)\right)$$
$$=\sum_i n_i g_\#\left(f_\#(\sigma_i)\right)=\sum_i n_i g\left(f(\sigma_i)\right)=\sum_i n_i(g\circ f)(\sigma_i)$$
$$=\sum_i n_i(g\circ f)_\#(\sigma_i)=(g\circ f)_\#\left(\sum_i n_i\sigma_i\right)=(g\circ f)_\#(\psi).$$

Therefore, for $\Omega=\omega\operatorname{im}\partial_{n+1}^X\in H_n(X)$ with $\omega\in\ker\partial_n^X$ we obtain by repeated application of the definitions of our functions,

$$(g_*\circ f_*)(\Omega)=g_*\left(f_*(\Omega)\right)=g_*\left(f_*\left(\omega\operatorname{im}\partial_{n+1}^X\right)\right)=g_*\left(f_\#(\omega)\operatorname{im}\partial_{n+1}^Y\right)$$
$$=g_\#\left(f_\#(\omega)\right)\operatorname{im}\partial_{n+1}^Z=(g\circ f)_\#(\omega)\operatorname{im}\partial_{n+1}^Z=(g\circ f)_*(\Omega). \quad \blacksquare$$

Now we introduce the concept of a homotopy between two maps. A reader unfamiliar with this may consult any introductory topology text to learn more. Two continuous maps $f:X\to Y$ and $g:X\to Y$ are said to be homotopic, denoted $f\sim g$, if there exists a continuous map $F:X\times[0,1]\to Y$ such that $F(x,0)=f(x)$ and $F(x,1)=g(x)$. For brevity we let $I=[0,1]$ below.

Example 6: The simplest example of a homotopy between two maps is the straight-line homotopy in $\mathbb{R}^2$. Let $f(x)=\sin x$ and $g(x)=x^2$. We know from basic calculus that both *f* and *g* are continuous functions. Define $F(x,t)=(1-t)f(x)+tg(x)$. It is easy to see that $F(x,0)=f(x)$ and $F(x,1)=g(x)$. Finally *F* is continuous since the sum of continuous functions is continuous. *F* is called the straight line homotopy because *F* joins the point $f(x)$ to the point $g(x)$ by means of a line segment in 2-space as *t* ranges from 0 to 1.

Theorem 10: Let $f : X \to Y$ and $g : X \to Y$ be homotopic maps. Let $p : U \to V \subseteq X$ and $q : W \subseteq Y \to Z$ be homeomorphisms. Further let $f(X) \subseteq W$ and $g(X) \subseteq W$. Then, $(f \circ p) \sim (g \circ p)$ and $(q \circ f) \sim (q \circ g)$.

Proof: Let $f : X \to Y$ and $g : X \to Y$ be homotopic maps. Let $p : U \to V \subseteq X$ and $q : W \subseteq Y \to Z$ be homeomorphisms with $f(X) \subseteq W$ and $g(X) \subseteq W$. Then there exists a continuous map $F : X \times I \to Y$ such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$. Define $G : U \times I \to Y$ by $G(u,t) = F(p(u),t)$. Since both *p* and *F* are continuous, their composition *G* is continuous. Further, $G(u,0) = F(p(u),0) = f(p(u))$ and $G(u,1) = F(p(u),1) = g(p(u))$. Thus, by definition, $(f \circ p) \sim (g \circ p)$. Similarly, the function $H : X \times I \to Z$ defined by $H(x,t) = q(F(x,t))$ is continuous with

$$H(x,0) = q(F(x,0)) = q(f(x))$$

and

$$H(x,1) = q(F(x,1)) = q(g(x)).$$

So, $(q \circ f) \sim (q \circ g)$. ■

We now supplement Theorem 10 with a theorem that will allow us to compute some homology groups very easily.

Theorem 11: Let $f : X \to Y$ and $g : X \to Y$ be homotopic maps. Then the induced maps on the homology groups are equal: $f_* = g_*$.

From an intuitive perspective this result is not particularly surprising. The homotopy equivalence between two maps *f* and *g* indicates essentially that the map *f* can be continuously molded into the map *g*. Therefore, it is only natural that given a simplex σ, the image of the simplex under *f* can be continuously molded into the image of the simplex under *g*. Therefore, if $f(\sigma) \in \ker \partial_n^Y$ it follows that $g(\sigma) \in \ker \partial_n^Y$ since a continuous deformation of $f(\sigma)$ into $g(\sigma)$ would suggest that $g(\sigma)$ has this same property as $f(\sigma)$. A similar intuitive argument can be made for elements in $\text{im}(\partial)$. Hence we intuit how *f* and *g* perform the same operation on the elements of the sets used to define homology groups and therefore induce the same map on homology groups.

We delay the proof of Theorem 11 until we have defined and explored what is called a prism map, the key idea that unlocks Theorem 11. We begin with a definition that seems esoteric; however, the motivation for it will be clarified in the proof of Theorem 11 below.

Definition 7: Let $F: X \times I \to Y$ be a homotopy for maps $f: X \to Y$ and $g: X \to Y$. Let $\sigma \in C_n(X)$. Let $\sigma \times 1: \Delta^n \times I \to X \times I$ where $I = [0,1]$ and 1 is the identity map on *I*. Finally, let $[v_0, v_1, \ldots, v_n]$ denote the simplex $\Delta^n \times \{0\}$ and let $[w_0, w_1, \ldots, w_n]$ denote the simplex $\Delta^n \times \{1\}$. Define the *prism map* $P: C_n(X) \to C_{n+1}(Y)$ by

$$P(\sigma) = \sum_i (-1)^i \left(F \circ (\sigma \times 1)\right)\Big|_{[v_0, v_1, \ldots, v_i, w_i, \ldots, w_n]}.$$

Note that $[v_0, v_1, \ldots, v_i, w_i, \ldots, w_n]$ is an $(n+1)$-simplex lying between $\Delta^n \times \{0\}$ and $\Delta^n \times \{1\}$.

Theorem 12: The prism map yields the following relationship: $g_\# = P\partial + \partial P + f_\#$.

Proof: Let $f: X \to Y$ and $g: X \to Y$. Let $F: X \times I \to Y$ be a homotopy from *f* to *g*. From the definition of the boundary map we see for generator $\omega \in C_n(X)$,

$$\begin{aligned}
\partial\left(P(\omega)\right) &= \partial\left(\sum_i (-1)^i \left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, v_i, w_i, \ldots, w_n]}\right) \\
&= \sum_j (-1)^j \sum_i (-1)^i \left(F \circ (\omega \times 1)\right)\Big|_{\underbrace{[v_0, v_1, \ldots, v_i, w_i, \ldots, w_n]}_{\text{with the } j\text{th term removed}}} \\
&= \sum_{j \le i} (-1)^j (-1)^i \left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, \hat{v}_j, \ldots, v_i, w_i, \ldots, w_n]} \\
&\qquad + \sum_{j \ge i} (-1)^{j+1} (-1)^i \left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, v_i, w_i, \ldots, \hat{w}_j, \ldots, w_n]},
\end{aligned}$$

where we have split the sum into two so that the first sum handles the removal of the *v* vertices and the second sum handles the removal of the *w* vertices. When $i = j$ the two sums reduce to

$$\begin{aligned}
&\sum_{j=i} (-1)^i (-1)^i \left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, \hat{v}_i, w_i, \ldots, w_n]} \\
&\qquad + \sum_{j=i} (-1)^{i+1} (-1)^i \left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, v_i, \hat{w}_i, \ldots, w_n]} \\
&= \sum_{i=j=0}^{n} \left(\left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, \hat{v}_i, w_i, \ldots, w_n]} - \left(F \circ (\omega \times 1)\right)\Big|_{[v_0, v_1, \ldots, v_i, \hat{w}_i, \ldots, w_n]}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(F\circ(\omega\times 1)\right)\Big|_{[\hat{v}_0, w_0, w_1,\dots,w_n]} - \left(F\circ(\omega\times 1)\right)\Big|_{[v_0, \hat{w}_0, w_1,\dots,w_n]} \\
&\quad + \left(F\circ(\omega\times 1)\right)\Big|_{[v_0, \hat{v}_1, w_1,\dots,w_n]} - \left(F\circ(\omega\times 1)\right)\Big|_{[v_0, v_1, \hat{w}_1,\dots,w_n]} \\
&\quad + \cdots + \left(F\circ(\omega\times 1)\right)\Big|_{[v_0, v_1,\dots,\hat{v}_n, w_n]} - \left(F\circ(\omega\times 1)\right)\Big|_{[v_0, v_1,\dots,v_n, \hat{w}_n]} \\
&= \left(F\circ(\omega\times 1)\right)\Big|_{[\hat{v}_0, w_0, w_1,\dots,w_n]} - \left(F\circ(\omega\times 1)\right)\Big|_{[v_0, v_1,\dots,v_n, \hat{w}_n]},
\end{aligned}
$$

where equality follows from the observation that the sum telescopes whereby only the first and last terms remain. The left term is simply $F\circ(\omega\times 1)$ restricted to the simplex $\Delta^n\times\{1\}$. Therefore,

$$\left(F\circ(\omega\times 1)\right)\Big|_{[\hat{v}_0, w_0, w_1,\dots,w_n]} = g(\omega) = g_\#(\omega).$$

Similarly, the right term is $-F\circ(\omega\times 1)$ restricted to the simplex $\Delta^n\times\{0\}$. Therefore,

$$-\left(F\circ(\omega\times 1)\right)\Big|_{[v_0, v_1,\dots,v_n, \hat{w}_n]} = -f(\omega) = -f_\#(\omega).$$

We also note that

$$
\begin{aligned}
P\left(\partial(\omega)\right) &= P\left(\sum_{i=0}^{n}(-1)^i\,\omega\Big|_{[v_0,\dots,\hat{v}_i,\dots,v_n]}\right) = \sum_j(-1)^j\sum_{i=0}^{n}(-1)^i\left(F\circ\left(\omega\Big|_{[v_0,\dots,\hat{v}_i,\dots,v_n]}\times 1\right)\right) \\
&= \sum_j(-1)^j\sum_{i=0}^{n}(-1)^i\left(F\circ(\omega\times 1)\right)\Big|_{[v_0,\dots,\hat{v}_i,\dots,v_j,w_j,\dots,w_n]},
\end{aligned}
$$

where we observe that with the v_i vertex removed in the simplex $\Delta^n\times\{0\}$ the corresponding w_i vertex in the simplex $\Delta^n\times\{1\}$ is also removed. However, by the definition of the prism map, only one of v_i or w_i may appear within the brackets at a time. Therefore, only one of these vertices appears with the "hat" notation. We also see there cannot be an instance where $j=i$ because the vertex indexed by *i* has already been removed. So, the above sum can be split into cases where $j<i$ and $j>i$:

$$\sum_{j<i}(-1)^i(-1)^j\left(F\circ(\omega\times 1)\right)\Big|_{[v_0,\dots,\hat{v}_j,\dots,v_i,w_i,\dots,w_n]} + \sum_{j>i}(-1)^{i+1}(-1)^j\left(F\circ(\omega\times 1)\right)\Big|_{[v_0,\dots,v_i,w_i,\dots,\hat{w}_j,\dots,w_n]}.$$

A close observation demonstrates that this is precisely $-\partial P(\omega)$ when $j\neq i$. Hence, it follows that $\partial P = -P\partial + g_\# - f_\#$, or equivalently, $g_\# = P\partial + \partial P + f_\#$. ■

With this relationship we return to the proof of Theorem 11.

Proof of Theorem 11: Let $f : X \to Y$ and $g : X \to Y$ be homotopic. Then there exists a function $F : X \times I \to Y$ such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$. With this in mind, if Δ^n is an *n*-dimensional simplex of the space *X*, then $\Delta^n \times I$ is homeomorphic to an $(n+1)$-dimensional simplex of $X \times I$, the domain of the homotopy *F*. Hence, it is only natural to consider the *n*-dimensional subsimplices of $\Delta^n \times I$.

Let $[v_0, v_1, \ldots, v_n]$ denote the simplex $\Delta^n \times \{0\}$ and $[w_0, w_1, \ldots, w_n]$ denote $\Delta^n \times \{1\}$. These are two *n*-dimensional simplices sitting in $(n+1)$-dimensional space. We are going to construct the $(n+1)$-dimensional simplices that occupy the space between these two simplices. We do this by starting with $[v_0, v_1, \ldots, v_n]$. We then remove the last vertex v_n and replace it with w_n in order to obtain $[v_0, v_1, \ldots, v_{n-1}, w_n]$. Effectively, we map the last vertex through the $(n+1)$-dimensional space between $\Delta^n \times \{0\}$ and $\Delta^n \times \{1\}$.

We construct another simplex by the same process to obtain $[v_0, v_1, \ldots, v_{n-2}, w_{n-1}, w_n]$. Continuing in this process we see that the $(n+1)$-dimensional simplices are of the form $[v_0, v_1, \ldots, v_i, w_i, \ldots, w_n]$. Given a generator $\sigma \in C_n(X)$, we consider

$$\sigma \times 1 : \Delta^n \times I \to X \times I$$

where 1 signifies the identity map on *I*.

We readily see the natural composition that can be formed: $\Delta^n \times I \xrightarrow{\sigma \times 1} X \times I \xrightarrow{F} Y$. This analysis yields the motivation to define the *prism map* as we did above.

Now let $\Omega \in H_n(X)$ where $\Omega = \omega \operatorname{im} \partial_{n+1}^X$ for some $\omega \in \ker \partial_n^X$. Since $\omega \in \ker \partial_n^X$ we have $P\left(\partial_n^X(\omega)\right) = P(0) = 0$. By Theorem 12,

$$\begin{aligned}
g_*(\Omega) &= g_\#(\omega) \operatorname{im}\left(\partial_{n+1}^Y\right) \\
&= \left(P\left(\partial_n^X(\omega)\right) + \partial_{n+1}^Y\left(P(\omega)\right) + f_\#(\omega)\right) \operatorname{im}\left(\partial_{n+1}^Y\right) \\
&= \left(\partial_{n+1}^Y\left(P(\omega)\right) + f_\#(\omega)\right) \operatorname{im}\left(\partial_{n+1}^Y\right) \\
&= \left(\partial_{n+1}^Y\left(P(\omega)\right) \operatorname{im}\left(\partial_{n+1}^Y\right)\right) + \left(f_\#(\omega) \operatorname{im}\left(\partial_{n+1}^Y\right)\right) \\
&= \operatorname{im}\left(\partial_{n+1}^Y\right) + \left(f_\#(\omega) \operatorname{im}\left(\partial_{n+1}^Y\right)\right) \\
&= \left(f_\#(\omega) \operatorname{im}\left(\partial_{n+1}^Y\right)\right) \\
&= f_*(\Omega),
\end{aligned}$$

where the third to last equation follows from the fact that $\partial_{n+1}^{Y}\left(P(\omega)\right)\in \operatorname{im}\left(\partial_{n+1}^{Y}\right)$ and the second to last equation follows from the fact that $\operatorname{im}\left(\partial_{n+1}^{Y}\right)$ acts as the identity for the group operation which we have denoted with the "+" symbol. Thus, the homomorphisms on homology groups induced by homotopy equivalent maps are the same. ■

For our purposes we need to compute the homology groups of $\mathbb{R}^n$ and S^n. With Theorem 11 we may easily compute homology groups of the former.

Theorem 13: For $i>0$, $H_i\left(\mathbb{R}^n\right)=0$.

Proof: Let $f:\mathbb{R}^n\to\mathbb{R}^n$ be the identity map on $\mathbb{R}^n$ and let $g:\mathbb{R}^n\to\mathbb{R}^n$ be a constant map sending each point of $\mathbb{R}^n$ to a fixed point *p*.

Clearly, $F:\mathbb{R}^n\times I\to\mathbb{R}^n$ by $F(x,t)=(1-t)f(x)+tg(x)$ is a homotopy. By Theorem 11, $f_*=g_*$. Let $\Omega\in H_i\left(\mathbb{R}^n\right)$ such that $\Omega=\omega\operatorname{im}\partial_{i+1}$ for $\omega=\sum_j n_j\sigma_j\in\ker\partial_i$. Then

$$f_*(\Omega)=f_\#(\omega)\operatorname{im}\partial_{i+1}=f_\#\left(\sum_j n_j\sigma_j\right)\operatorname{im}\partial_{i+1}=\left(\sum_j n_j f_\#\left(\sigma_j\right)\right)\operatorname{im}\partial_{i+1}$$
$$=\left(\sum_j n_j f\left(\sigma_j\right)\right)\operatorname{im}\partial_{i+1}=\left(\sum_j n_j\sigma_j\right)\operatorname{im}\partial_{i+1}=\omega\operatorname{im}\partial_{i+1}=\Omega\ .$$

That is, f_* is the identity isomorphism on $H_i\left(\mathbb{R}^n\right)$.

Since *g* is the constant map on $\mathbb{R}^n$, *g* induces a homomorphism $g_*:H_i\left(\mathbb{R}^n\right)\to H_i\left(\{p\}\right)$ where $\{p\}$ is just the space of a single point. However, since $g_*=f_*$ and f_* is an isomorphism, it follows that g_* is also an isomorphism. Thus, $H_i\left(\mathbb{R}^n\right)\cong H_i\left(\{p\}\right)$. We can more easily compute the homology groups of a single point.

For the space $\{p\}$, the maps $\sigma_\alpha^i:\Delta^i\to\{p\}$ are unique for all $i\geq 0$ since, of course, there can be only one map per *i* that takes the space Δ^i and sends it to the point *p*. Therefore, $C_i\left(\{p\}\right)=\left\langle\sigma^i\right\rangle\cong\mathbb{Z}$ for all $i\geq 0$. Observe that for $\sigma^i\in C_i\left(\{p\}\right)$ with $i>0$,

$$\partial_i\left(\sigma^i\right)=\sum_{j=0}^{i}(-1)^j\,\sigma^i\Big|_{\left[v_0,v_1,\ldots,\hat{v}_j,\ldots,v_i\right]}\ .$$

Since the map σ^i is constant,

$$\sum_{j=0}^{i}(-1)^j \sigma^i\big|_{[v_0,v_1,\ldots,\hat{v}_j,\ldots,v_i]} = \begin{cases} \sigma^i\big|_{[v_0,v_1,\ldots,\hat{v}_j,\ldots,v_i]} & \text{if } i \text{ is even} \\ 0 & \text{if } i \text{ is odd} \end{cases}$$

This implies that $\operatorname{im}\partial_i = 0$ when i is odd and $\operatorname{im}\partial_i \cong C_{i-1}(\{p\}) \cong \mathbb{Z}$ when i is even and $i > 0$. Therefore, for i even and $i > 0$ we have the diagram:

$$\cdots \to \mathbb{Z} \xrightarrow{\partial_{i+2}=\mathrm{id}} \mathbb{Z} \xrightarrow{\partial_{i+1}=0} \mathbb{Z} \xrightarrow{\partial_i=\mathrm{id}} \mathbb{Z} \xrightarrow{\partial_{i-1}=0} \mathbb{Z} \xrightarrow{\partial_{i-2}=\mathrm{id}} \cdots,$$

where id represents the identity isomorphism. Thus, for $i > 0$ and i even,

$$H_i(\{p\}) = \ker\partial_i / \operatorname{im}\partial_{i+1} = 0/0 = 0.$$

Similarly, for i odd, we have the diagram:

$$\cdots \to \mathbb{Z} \xrightarrow{\partial_{i+2}=0} \mathbb{Z} \xrightarrow{\partial_{i+1}=\mathrm{id}} \mathbb{Z} \xrightarrow{\partial_i=0} \mathbb{Z} \xrightarrow{\partial_{i-1}=\mathrm{id}} \mathbb{Z} \xrightarrow{\partial_{i-2}=0} \cdots,$$

where id, again, represents the identity isomorphism. Thus, for i odd,

$$H_i(\{p\}) = \ker\partial_i / \operatorname{im}\partial_{i+1} = \mathbb{Z}/\mathbb{Z} = 0.$$

Thus,

$$H_i(\mathbb{R}^n) \cong H_i(\{p\}) = 0 \text{ for } i > 0. \quad \blacksquare$$

We now proceed to develop the machinery to compute the homology group of the n-dimensional sphere.

Definition 8: For a space X, the suspension of X, denoted ΣX , is the quotient space formed from $X \times I$ by collapsing $X \times \{0\}$ and $X \times \{1\}$ to separate points, the former called the "south pole" and the latter called the "north pole."

When the space under consideration is the n-dimensional sphere S^n with $n > 0$, it is easy to see that $\Sigma S^n = S^{n+1}$.

Definition 9: Let $f : X \to Y$ be continuous. We define $\Sigma f : \Sigma X \to \Sigma Y$, the suspension map from the suspension of X to the suspension of Y , as the quotient map of $f \times i : X \times I \to Y \times I$ where I is the closed unit interval $[0,1]$ and i is the identity map on I.

It is worth expanding on the definition of the suspension map. We refer to [5] for the source of this discussion. We may think of the suspension map as taking the point

$(x,t) \in X \times I$ to the point $(f(x),t) \in Y \times I$ and then performing the necessary contraction of $Y \times \{1\}$ and $Y \times \{0\}$ into the north and south poles, respectively. Since we will consider the suspension map in the case where $X = Y = S^n$, we will give a more explicit example. Let $f : S^n \to S^n$ be continuous, let $\Sigma f : S^{n+1} \to S^{n+1}$ be the suspension of *f*, and let $s = (x_0, x_1, \dots, x_{n+1}) \in S^{n+1}$. Then the suspension map in this case is

$$\Sigma f(s) = \Sigma f((x_0, x_1, \dots, x_{n+1})) = \left(\left(\sqrt{1-x_{n+1}^2}\right) f\left(\frac{x_0}{\sqrt{1-x_{n+1}^2}}, \frac{x_1}{\sqrt{1-x_{n+1}^2}}, \dots, \frac{x_n}{\sqrt{1-x_{n+1}^2}} \right), x_{n+1} \right).$$

The process of taking the quotient is captured by dividing the coordinates $x_0, x_1, \dots, x_n$ by $\sqrt{1-x_{n+1}^2}$. Since the domain of *f* is S^n we cannot simply write $f(x_0, x_1, \dots, x_n)$ since the point $(x_0, x_1, \dots, x_n)$ may not be an element of S^n. Therefore, dividing by $\sqrt{1-x_{n+1}^2}$ ensures that the point $\left(\frac{x_0}{\sqrt{1-x_{n+1}^2}}, \frac{x_1}{\sqrt{1-x_{n+1}^2}}, \dots, \frac{x_n}{\sqrt{1-x_{n+1}^2}} \right)$ is in S^n.

To get an intuitive feel for why the suspension map is important, observe that under the suspension operation a typical *n*-simplex is homeomorphic to an $(n+1)$-simplex. This should be clear from the definition of suspension. For the sake of illustration, begin with a point, a 0-simplex. Under the suspension operation, this point becomes a line, a 1-simplex. Under the suspension operation again, the line becomes something like a solid ellipse, two of whose opposite points are pinched down. Such a shape is, indeed, homeomorphic to a 2-simplex. Continuing with this kind of intuitive induction, we are convinced.

We can make these concepts more precise with the following theorem.

Theorem 14: For $n \geq 0$ $H_n(X) \cong H_{n+1}(\Sigma X)$.

Proof: The suspension map can be conceived as a continuous map $\Sigma : X \to \Sigma X$. In fact, we can see that Σ is a bijection. Let $x \in X$. Then, the suspension of a point *x* is a set of points homeomorphic to the set of points $\{(x,t) | t \in (0,1)\}$ and with $(x,0)$ identified with the south pole and $(x,1)$ identified with the north pole of ΣX.

Suppose for $x, y \in X$ that $\Sigma x = \Sigma y$. Then, it follows that

$$\{(x,t) | t \in (0,1)\} \cong \{(y,s) | s \in (0,1)\},$$

and $(x,0)=(y,0)$ and $(x,1)=(y,1)$.

Therefore, for any (x,t) there exist some $y\in X$ such that (x,t) can be identified with (y,t). Equality holds if and only if $x=y$. Thus, Σ is injective.

It is rather trivial that Σ is surjective. For some $z\in\Sigma X$, z is mapped to by some (w,t) for some $w\in X$ and $t\in[0,1]$ by the definition of suspension. Thus, Σ is bijective.

Therefore, Σ induces the isomorphism $\Sigma_\#:C_n(X)\to C_{n+1}(\Sigma X)$ defined on the generators $\omega\in C_n(X)$ by $\Sigma_\#(\omega)=\Sigma\omega$. Further, Σ also induces the isomorphism

$$\Sigma_*:H_n(X)\to H_{n+1}(\Sigma X)$$

defined by

$$\Sigma_*(\Omega)=\left(\Sigma_\#(\omega)\right)\mathrm{im}\,\partial_{n+2}^{\Sigma X}=(\Sigma\omega)\mathrm{im}\,\partial_{n+2}^{\Sigma X},$$

where $\Omega=\omega\,\mathrm{im}\,\partial_{n+1}^{X}$ with $\omega\in\ker\partial_n^X$. We note explicitly that if $\omega\in\ker\partial_n^X$ then $\Sigma\omega\in\ker\partial_{n+1}^{\Sigma X}$ by the commutativity of boundary maps with the induced map $\Sigma_\#$ analogous to the commutativity of the boundary maps with the map $f_\#$ in Theorem 7. ■

When $X=S^n$ we derive $H_n(S^n)=H_{n+1}(S^{n+1})$ for all $n>0$. And so, by a simple induction argument

$$H_1(S^1)\cong H_2(S^2)\cong H_3(S^3)\cong\cdots\cong H_n(S^n)\cong\cdots.$$

Theorem 15: Let S^1 be the circle in the plane. Then $H_1(S^1)=\mathbb{Z}$.

Proof: We note that the 2-simplex is homeomorphic to the circle and so all embeddings of the 2-simplex into the circle are homeomorphisms. That is, $\sigma^2:\Delta^2\to S^1$ is a homeomorphism. The boundary of all such embeddings must be 0 since the boundary of the circle is 0. Therefore $\mathrm{im}\,\partial_2=0$. Now, let $\sigma^1:\Delta^1\to S^1$ be an embedding of a 1-simplex in the circle. Clearly, the embedding must be an arc of the circle. To evaluate $\ker\partial_1$ we simply ask what sort of arcs on the circle have a boundary of 0. That is, which arcs on the circle begin and end at the same point? That's easy: the arcs that wrap around the entire circle either clockwise or counter-clockwise. The arcs are distinguishable by the number of times they wrap around the circle. Therefore the congruence classes of the arcs are the integers. Thus $\ker\partial_1=\mathbb{Z}$. Hence, $H_1(S^1)=\ker\partial_1/\mathrm{im}\,\partial_2=\mathbb{Z}$ as desired. ■

Corollary 2: $H_n\left(S^n\right)=\mathbb{Z}$ for all $n>0$.

Proof : This is a direct corollary from Theorem 14 and Theorem 15. ■

Now it is a very important result that the *n*th homology group of an *n*-dimensional sphere is the group of integers. In particular, this means that the induced homomorphism

$$f_*^n : H_n\left(S^n\right)\to H_n\left(S^n\right)$$

is a homomorphism from the integers to the integers. The following theorem observes that such a homomorphism must be simple multiplication.

Theorem 16: Let $\varphi:\mathbb{Z}\to\mathbb{Z}$ be a group homomorphism. Then, $\varphi(x)=nx$ for some integer *n*.

Proof: Let $\varphi:\mathbb{Z}\to\mathbb{Z}$ be a group homomorphism. Then, $\varphi(1)=n$ for some integer *n*. Therefore, for any $m\in\mathbb{Z}$, $\varphi(m)=\varphi(m\cdot 1)=m\varphi(1)=nm$ as desired. ■

This theorem motivates the following definition attributable to Brouwer.

Definition 10: Let $f:S^n\to S^n$ be continuous. Then, the induced map

$$f_*^n : H_n\left(S^n\right)\to H_n\left(S^n\right)$$

is a homomorphism from $\mathbb{Z}$ to $\mathbb{Z}$. By the preceding theorem, $f_*^n(\Omega)=n\Omega$ for some integer *n*. This integer *n* is defined as the *topological degree of f*, denoted $\deg(f)=n$.

Since we will be dealing with polynomials over the quaternions, we are interested in extensions of functions of the form x^m to the 5-dimensional sphere S^4. Given a map $x^m:\mathbb{H}\to\mathbb{H}$, we want to know what the topological degree of $\left(p^{-1}\circ x^m\circ p\right):S^4\to S^4$ is, where $p:S^4\to\mathbb{R}^4\cup\{\infty\}$ and $p^{-1}:\mathbb{R}^4\cup\{\infty\}\to S^4$ are, respectively, the stereographic projection and compactification maps with $\mathbb{R}^4$ regarded as the space $\mathbb{H}$. We make explicit here that for the stereographic projection on the 5-sphere $p\left((0,0,0,0,1)\right)=\infty$ and for the compactification map of 4-dimensional Euclidean space into the 5-dimensional sphere $p^{-1}(\infty)=(0,0,0,0,1)$. Going forward, for the sake of concision, we will not explicitly state these two relationships.

Unsurprisingly, the topological degree of the composition map $\left(p^{-1}\circ x^m\circ p\right)$ and the algebraic degree of x^m coincide. This fact we now aim to prove.

Theorem 17: Let $f : S^n \to S^n$. Then $\deg(f) = \deg(\Sigma f)$.

Proof: Let $f : S^n \to S^n$. Then $\Sigma f : S^{n+1} \to S^{n+1}$. Let $\Sigma : S^n \to S^{n+1}$. We can easily see that $\Sigma \circ f = (\Sigma f) \circ \Sigma$. For $x \in S^n$, the suspension of the point *x* is a set of points homeomorphic to the set $\{(x,t) | t \in (0,1)\}$ with $(x,0)$ identified with the south pole of S^{n+1} and $(x,1)$ identified with the north pole of S^{n+1}. Then, from the definition of the suspension map, $(\Sigma f)(\Sigma x)$ is homeomorphic to the set $\{(f(x),t) | t \in (0,1)\}$ with $(f(x),0)$ identified with the south pole and $(f(x),1)$ identified with the north pole.

Similarly, for $x \in S^n$, $\Sigma(f(x))$ is homeomorphic to the set $\{(f(x),t) | t \in (0,1)\}$ with $(f(x),0)$ identified with the south pole and $(f(x),1)$ identified with the north pole.

Since *x* is an arbitrary point, it follows that $\Sigma \circ f = (\Sigma f) \circ \Sigma$. Thus, by the definitions of the induced maps $f_\#$, $\Sigma_\#$, and $(\Sigma f)_\#$, it follows that $\Sigma_\# \circ f_\# = (\Sigma f)_\# \circ \Sigma_\#$. Therefore, it is also true that $\Sigma_* \circ f_* = (\Sigma f)_* \circ \Sigma_*$.

By Theorem 16 there exist integers *n* and *m* such that $f_*(\Omega) = n\Omega$ and $(\Sigma f)_*(\Psi) = m\Psi$.

For $\Omega \in H_n(S^n)$, we have

$$\begin{aligned} n\Sigma_*(\Omega) &= \Sigma_*(n\Omega) \\ &= \Sigma_*(f_*(\Omega)) \\ &= (\Sigma f)_*(\Sigma_*(\Omega)) \\ &= m\Sigma_*(\Omega) \end{aligned}$$

and thus $m = n$. ■

Theorem 18: Let $f : S^1 \to S^1$ be defined by $f(z) = z^k$ for some $k \in \mathbb{N}$. Then $\deg(f) = k$.

Proof: Define $f : S^1 \to S^1$ by $f(z) = z^k$ for some positive integer *k*. Consider an arbitrarily long arc on S^1, which we denote as *A*. The function *f* stretches *A* by a factor of *k*.

So, $f_\# : C_1\left(S^1\right) \to C_1\left(S^1\right)$ takes 1-simplices and stretches them by a factor of *k*. For an individual generator this translates to $\sigma \mapsto k\sigma$. For finite sums of generators we have

$$f_\#\left(\sum_i n_i\sigma_i\right) = \sum_i n_i f_\#\left(\sigma_i\right) = \sum_i n_i k\sigma_i = k\sum_i n_i\sigma_i$$

so that $f_\#$ is multiplication by *k*. Therefore, the induced homomorphism on homology groups is also multiplication by *k*. For $\Omega \in H_1\left(S^1\right)$,

$$f_*\left(\Omega\right) = f_\#\left(\omega\right)\operatorname{im}\partial_2^Y = k\omega\operatorname{im}\partial_2^Y = k\Omega .$$

Thus $\deg\left(f\right) = k$. ■

Definition 11: Let $k \in \mathbb{N}$. Define $f : S^1 \to S^1$ by $f\left(z\right) = z^k$. We define $z^k : S^n \to S^n$ to be the $\left(n-1\right)$-st iterated suspension of *f*. That is $z^k = \left(\Sigma^{n-1} f\right) : S^n \to S^n$. Then by Theorem 17 and Theorem 18, $\deg\left(z^k\right) = \deg\left(\Sigma^{n-1} f\right) = \deg\left(f\right) = k$.

Theorem 19: Let $f : \mathbb{H} \to \mathbb{H}$ be defined by $f\left(q\right) = q^k$. Let $p : S^4 \to \mathbb{H} \cup \left\{\infty\right\}$ be the stereographic projection of the 5-sphere into the quaternions and let $p^{-1} : \mathbb{H} \cup \left\{\infty\right\} \to S^4$ be the one-point compactification of the quaternions into the 5-sphere. Then $\deg\left(p^{-1} \circ f \circ p\right) = k$.

The result is not particularly surprising, but it is important enough that we should develop a rigorous proof of it. Before we build up the machinery to prove this theorem, we first give a motivating example.

Example 7: Let $z^k : S^2 \to S^2$ be the suspension of $z^k : S^1 \to S^1$. Let $p : S^2 \to \mathbb{R}^2 \cup \left\{\infty\right\}$ be the stereographic projection of the 2-sphere into the plane, where $p\left(\left(0,0,1\right)\right) = \infty$. Let $p^{-1} : \mathbb{R}^2 \cup \left\{\infty\right\} \to S^2$ be the one-point compactification of the *xy*-plane into the 2-sphere where $p^{-1}\left(\infty\right) = \left(0,0,1\right)$, and let $x^k : \mathbb{R}^2 \to \mathbb{R}^2$ be multiplication of ordered pairs.

Let *C* denote a cross section of S^2 obtained from intersecting the 2-sphere with plane P_1 that lies between $-1 < z < 1$ and that is parallel to the *xy*-plane. Then *C* is a circle of radius $0 < r < 1$. The definition of the suspension map can be broken down into three operations on the circle *C*. The suspension map first maps *C* to a circle in P_1 with radius

1. Call this circle *D*. Now each point on *D* can be written as $(\cos\theta, \sin\theta)$ for some angle $\theta \in [0, 2\pi)$. And so the suspension map maps the points of *D* accordingly

$$(\cos\theta, \sin\theta) \mapsto (\cos k\theta, \sin k\theta).$$

Call the image of *D* under the multiplication operation of the suspension map E. Finally, the suspension map retracts the circle *E* onto a circle *F* of the same radius as *C*. In short, then, the suspension map sends *C* in plane P_1 to *F* also in P_1 where *F* is obtained from *C* by rotating the points of *C* by the multiple *k*.

Now, consider *C* under the composition $p^{-1} \circ x^k \circ p$. The image $p(C)$ is a circle, *G*, of radius *s* in the *xy*-plane. The points of *G* may be expressed as $s(\cos\varphi, \sin\varphi)$ for $\varphi \in [0, 2\pi)$. The map x^k maps the points of *G* by

$$s(\cos\varphi, \sin\varphi) \mapsto s^k (\cos k\varphi, \sin k\varphi).$$

We let *H* denote the circle obtained from $x^k(G)$. Finally, p^{-1} sends *H* to a cross section *J* of the sphere that lies in a plane, P_2, that is parallel to the *xy*-plane. In other words, $J = (p^{-1} \circ x^k \circ p)(C)$ is a circle whose points are rotations by the multiple *k* of the points of *C* and that is translated from P_1 to P_2. In fact, it is intuitive that by translating P_1 to overlap with P_2 we may let *F* and *J* coincide. In other words, if we regard the translation of the planes as a homotopy (which it is) we see that $(p^{-1} \circ x^k \circ p)(C) \sim \Sigma(z^k)(C)$. With this example in mind, we may now develop these concepts more precisely in order to prove Theorem 19.

First of all, we note that it was quite helpful to write points in the plane in polar form. Therefore, we begin by developing an analogous expression for quaternions.

Let $t = (\alpha, \beta, \gamma, \delta) \in \mathbb{H} - \{0\}$. First we observe that when $\alpha \neq 0$

$$\frac{|\alpha|}{|t|} = \frac{|\alpha|}{\sqrt{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}} \leq \frac{|\alpha|}{\sqrt{\alpha^2}} = \frac{|\alpha|}{|\alpha|} = 1,$$

and $\frac{\alpha}{|t|} \leq 1$ trivially when $\alpha = 0$.

Therefore there exists a unique θ with $0 \leq \theta \leq \pi$ such that $\cos\theta = \frac{\alpha}{|t|}$.

We also note that θ satisfies

$$\sin^2\theta = 1-\cos^2\theta = 1-\left(\frac{\alpha}{|t|}\right)^2 = 1-\frac{\alpha^2}{|t|^2} = 1-\frac{\alpha^2}{\alpha^2+\beta^2+\gamma^2+\delta^2} = \frac{\beta^2+\gamma^2+\delta^2}{\alpha^2+\beta^2+\gamma^2+\delta^2}.$$

Since $0 \le \theta \le \pi$, $\sin\theta = \frac{\sqrt{\beta^2+\gamma^2+\delta^2}}{|t|}$.

Now if $\beta^2+\gamma^2+\delta^2 > 0$ we may define a vector

$$\bar{\xi} = \frac{1}{\sqrt{\beta^2+\gamma^2+\delta^2}}(0,\beta,\gamma,\delta),$$

and a simple computation shows that $\bar{\xi}^2 = -1$.

The expression $|t|\left(\cos\theta+\bar{\xi}\sin\theta\right)$ is easily seen to be the quaternion *t* since

$$\begin{aligned} |t|\left(\cos\theta+\bar{\xi}\sin\theta\right) &= |t|\left(\frac{\alpha}{|t|}+\left(\frac{1}{\sqrt{\beta^2+\gamma^2+\delta^2}}(0,\beta,\gamma,\delta)\right)\left(\frac{\sqrt{\beta^2+\gamma^2+\delta^2}}{|t|}\right)\right) \\ &= \alpha+(0,\beta,\gamma,\delta) \\ &= (\alpha,\beta,\gamma,\delta) \\ &= t. \end{aligned}$$

If $\beta^2+\gamma^2+\delta^2=0$, then $\sin\theta=0$. Therefore, $t=|t|\left(\cos\theta+\bar{\xi}\sin\theta\right)$ for any vector $\bar{\xi}$. For our purposes, if $\beta^2+\gamma^2+\delta^2=0$ we define $\bar{\xi}=\bar{0}$.

Suppose by way of mathematical induction that $\left(\cos\theta+\bar{\xi}\sin\theta\right)^k = \cos k\theta+\bar{\xi}\sin k\theta$ for all natural numbers up to *k*. It is trivial that the base case where $k=1$ holds. We recall the trigonometric identities

$$\cos(a+b) = \cos a\cos b - \sin a\sin b$$

and

$$\sin(a+b) = \sin a\cos b + \sin b\cos a,$$

and derive from the induction hypothesis that

$$\begin{aligned}\left(\cos\theta+\overline{\xi}\sin\theta\right)^n &= \left(\cos\theta+\overline{\xi}\sin\theta\right)^{n-1}\left(\cos\theta+\overline{\xi}\sin\theta\right)\\ &= \left(\cos\left((n-1)\theta\right)+\overline{\xi}\sin\left((n-1)\theta\right)\right)\left(\cos\theta+\overline{\xi}\sin\theta\right)\\ &= \cos\left((n-1)\theta\right)\cos\theta+\overline{\xi}\cos\left((n-1)\theta\right)\sin\theta\\ &\quad +\overline{\xi}\sin\left((n-1)\theta\right)\cos\theta-\sin\left((n-1)\theta\right)\sin\theta\\ &= \left(\cos\left((n-1)\theta\right)\cos\theta-\sin\left((n-1)\theta\right)\sin\theta\right)\\ &\quad +\overline{\xi}\left(\cos\left((n-1)\theta\right)\sin\theta+\sin\left((n-1)\theta\right)\cos\theta\right)\\ &= \cos n\theta+\overline{\xi}\sin n\theta.\end{aligned}$$

Thus in this discussion, we have shown that a nonzero quaternion may be expressed in terms of a unique angle $\theta\in[0,\pi]$ and vector $\overline{\xi}$. Furthermore, the expression $\cos\theta+\overline{\xi}\sin\theta$ defined above has similar attributes to the polar expression of a complex number in $\mathbb{C}$. Therefore, we make the following definition.

Definition 12: Let $t=(\alpha,\beta,\gamma,\delta)\in\mathbb{H}-\{0\}$. Define $\theta\in[0,\pi]$ such that

$$\cos\theta=\frac{\alpha}{|t|} \text{ and } \sin\theta=\frac{\sqrt{\beta^2+\gamma^2+\delta^2}}{|t|}.$$

If $\beta^2+\gamma^2+\delta^2>0$ let $\overline{\xi}=\frac{1}{\sqrt{\beta^2+\gamma^2+\delta^2}}(0,\beta,\gamma,\delta)$. If $\beta^2+\gamma^2+\delta^2=0$, let $\overline{\xi}=0$. Then

$$t=|t|\left(\cos\theta+\overline{\xi}\sin\theta\right).$$

We shall call the term $|t|\left(\cos\theta+\overline{\xi}\sin\theta\right)$ the *hyper-polar* form of *t*. We call the individual expressions

$$\cos\theta=\frac{\alpha}{|t|} \text{ and } \sin\theta=\frac{\sqrt{\beta^2+\gamma^2+\delta^2}}{|t|}$$

the *hyper-polar expressions* of *t*.

We may now use the hyper-polar expressions of quaternions to prove the following theorem which will be the key to proving Theorem 19.

Theorem 20: Let $k\in\mathbb{N}$. Let $p:S^4\to\mathbb{H}\cup\{\infty\}$ be the stereographic projection of the 5-dimensional sphere into 4-space which we regard as the space of quaternions. Let $\Sigma^3 s:S^4\to S^4$ be the 3rd iterated suspension of the map $s:S^1\to S^1$ defined by $s(z)=z^k$.

Let $q^k : \mathbb{H} \to \mathbb{H}$ be exponentiation of quaternions to the *k*th power. Then $\left(q^k \circ p\right) \sim \left(p \circ \Sigma^3 s\right)$.

Proof: Let $k \in \mathbb{N}$. Let $p : S^4 \to \mathbb{H} \cup \{\infty\}$ be the stereographic projection of the 5-dimensional sphere into 4-space which we regard as the space of quaternions. Let $\Sigma^3 s : S^4 \to S^4$ be the 3rd iterated suspension of the map $s : S^1 \to S^1$ defined by $s(z) = z^k$ and let $q^k : \mathbb{H} \to \mathbb{H}$ be exponentiation of quaternions to the *k*th power.

Let $(a,b,c,d,e) \in S^4 - \{(0,0,0,0,1)\}$. Then

$$\left(q^k \circ p\right)\left((a,b,c,d,e)\right) = \left(q^k\right)\left(p\left((a,b,c,d,e)\right)\right) = \left(q^k\right)\left(\left(\frac{a}{1-e}, \frac{b}{1-e}, \frac{c}{1-e}, \frac{d}{1-e}\right)\right).$$

We may express $\left(\frac{a}{1-e}, \frac{b}{1-e}, \frac{c}{1-e}, \frac{d}{1-e}\right) \in \mathbb{H}$ as

$$\left(\frac{\sqrt{a^2+b^2+c^2+d^2}}{1-e}\right)\left(\cos\theta + \bar{\xi}\sin\theta\right),$$

for $0 \le \theta \le \pi$ satisfying $\cos\theta = \frac{a}{\sqrt{a^2+b^2+c^2+d^2}}$ and $\sin\theta = \frac{\sqrt{b^2+c^2+d^2}}{\sqrt{a^2+b^2+c^2+d^2}}$.

Further , if $\sqrt{b^2+c^2+d^2} > 0$, we have the unit vector $\bar{\xi} = \frac{(0,b,c,d)}{\sqrt{b^2+c^2+d^2}}$. Otherwise, we set $\bar{\xi} = 0$. Then,

$$\begin{aligned}
&\left(q^k\right)\left(\left(\frac{a}{1-e}, \frac{b}{1-e}, \frac{c}{1-e}, \frac{d}{1-e}\right)\right) \\
&= \left(q^k\right)\left(\left(\frac{\sqrt{a^2+b^2+c^2+d^2}}{1-e}\right)\left(\cos\theta + \bar{\xi}\sin\theta\right)\right) \\
&= \left(\frac{\sqrt{a^2+b^2+c^2+d^2}}{1-e}\right)^k \left(\cos k\theta + \bar{\xi}\sin k\theta\right) \\
&= \left(\frac{\sqrt{a^2+b^2+c^2+d^2}}{1-e}\right)^k \left(\cos k\theta, \frac{b\sin k\theta}{\sqrt{b^2+c^2+d^2}}, \frac{c\sin k\theta}{\sqrt{b^2+c^2+d^2}}, \frac{d\sin k\theta}{\sqrt{b^2+c^2+d^2}}\right).
\end{aligned}$$

Define $F : \left(S^4 - \{(0,0,0,0,1)\}\right) \times I \to \mathbb{H}$, where I is the unit interval $[0,1]$, by

$$F(a,b,c,d,e,t) = \frac{\left(\cos k\theta, \frac{((1-t)b+t)\sin k\theta}{\left(t+(1-t)\sqrt{b^2+c^2+d^2}\right)}, \frac{c(1-t)\sin k\theta}{\sqrt{b^2+c^2+d^2}}, \frac{d(1-t)\sin k\theta}{\sqrt{b^2+c^2+d^2}}\right)}{\frac{1}{\left(\frac{\sqrt{a^2+b^2+c^2(1-t)+d^2(1-t)}}{1-e}\right)^k}}.$$

Then F is continuous since each of the component functions is continuous as a product of continuous functions. In particular, note that $t+(1-t)\sqrt{b^2+c^2+d^2} \neq 0$ for any $t \in [0,1]$, unless, of course, $\sqrt{b^2+c^2+d^2} = 0$ in which case $\bar{\xi} = 0$ and the function F reduces to

$$F(a,0,0,0,e,t) = \left(\frac{|a|}{1-e}\right)^k (\cos k\theta, 0, 0, 0).$$

Now, when $t = 0$ we see

$$\begin{aligned} F(a,b,c,d,e,0) &= \frac{\left(\cos k\theta, \frac{b\sin k\theta}{\sqrt{b^2+c^2+d^2}}, \frac{c\sin k\theta}{\sqrt{b^2+c^2+d^2}}, \frac{d\sin k\theta}{\sqrt{b^2+c^2+d^2}}\right)}{\frac{1}{\left(\frac{\sqrt{a^2+b^2+c^2+d^2}}{1-e}\right)^k}} \\ &= \left(q^k \circ p\right)\left((a,b,c,d,e)\right). \end{aligned}$$

When $t = 1$,

$$F(a,b,c,d,e,1) = \left(\frac{\sqrt{a^2+b^2}}{1-e}\right)^k (\cos k\theta, \sin k\theta, 0, 0).$$

Define $u : S^4 - \{(0,0,0,0,1)\} \to \mathbb{R}^4$ by

$$u(a,b,c,d,e) = \left(\frac{\sqrt{a^2+b^2}}{1-e}\right)^k (\cos k\theta, \sin k\theta, 0, 0)$$

where θ is determined by

$$\cos\theta = \frac{a}{\sqrt{a^2+b^2+c^2+d^2}} \text{ and } \sin\theta = \frac{\sqrt{b^2+c^2+d^2}}{\sqrt{a^2+b^2+c^2+d^2}} .$$

Then $\left(q^k \circ p\right) \sim u$.

Next we turn to the 3rd iterated suspension map.

Again, let $(a,b,c,d,e) \in S^4 - \{(0,0,0,0,1)\}$. Then,

$$\begin{aligned}
&\Sigma^3 s(a,b,c,d,e) \\
&= \left(\sqrt{1-e^2}\,\Sigma^2 s\left(\frac{a}{\sqrt{1-e^2}},\frac{b}{\sqrt{1-e^2}},\frac{c}{\sqrt{1-e^2}},\frac{d}{\sqrt{1-e^2}}\right),e\right) \\
&= \left(\sqrt{1-\frac{d^2}{1-e^2}}\sqrt{1-e^2}\,\Sigma s\left(\frac{a}{\sqrt{1-\frac{d^2}{1-e^2}}\sqrt{1-e^2}},\frac{b}{\sqrt{1-\frac{d^2}{1-e^2}}\sqrt{1-e^2}},\frac{c}{\sqrt{1-\frac{d^2}{1-e^2}}\sqrt{1-e^2}}\right),d,e\right) \\
&= \left(\sqrt{1-d^2-e^2}\,\Sigma s\left(\frac{a}{\sqrt{1-d^2-e^2}},\frac{b}{\sqrt{1-d^2-e^2}},\frac{c}{\sqrt{1-d^2-e^2}}\right),d,e\right).
\end{aligned}$$

Since $\sqrt{1-\frac{c^2}{1-d^2-e^2}}\sqrt{1-d^2-e^2} = \sqrt{1-d^2-e^2-c^2} = \sqrt{a^2+b^2}$, we derive

$$\begin{aligned}
&\left(\sqrt{1-d^2-e^2}\,\Sigma s\left(\frac{a}{\sqrt{1-d^2-e^2}},\frac{b}{\sqrt{1-d^2-e^2}},\frac{c}{\sqrt{1-d^2-e^2}}\right),d,e\right) \\
&\quad = \left(\sqrt{a^2+b^2}\,s\left(\frac{a}{\sqrt{a^2+b^2}},\frac{b}{\sqrt{a^2+b^2}}\right),c,d,e\right).
\end{aligned}$$

Now, if $a^2+b^2>0$, then $\left(\frac{a}{\sqrt{a^2+b^2}},\frac{b}{\sqrt{a^2+b^2}}\right) \in S^1$ and so there exists a $\tau \in [0,\pi]$ such that

$$\cos\tau = \frac{a}{\sqrt{a^2+b^2}} \text{ and } \sin\tau = \frac{b}{\sqrt{a^2+b^2}} .$$

The map s sends

$$(\cos\tau,\sin\tau) \mapsto (\cos k\tau,\sin k\tau) .$$

Therefore

$$\Sigma^3 s(a,b,c,d,e) = \left(\sqrt{a^2+b^2}\cos k\tau, \sqrt{a^2+b^2}\sin k\tau, c, d, e\right),$$

where $\cos k\tau$ and $\sin k\tau$ are determined by $\cos\tau = \dfrac{a}{\sqrt{a^2+b^2}}$ and $\sin\tau = \dfrac{b}{\sqrt{a^2+b^2}}$ when $a^2+b^2>0$. In the case when $a^2+b^2=0$, we set $\tau=0$.

Composing this result with *p* we obtain

$$\begin{aligned}\left(p\circ\Sigma^3 s\right)\left((a,b,c,d,e)\right) &= p\left(\left(\sqrt{a^2+b^2}\cos k\tau, \sqrt{a^2+b^2}\sin k\tau, c, d, e\right)\right)\\ &= \left(\frac{\sqrt{a^2+b^2}}{1-e}\cos k\tau, \frac{\sqrt{a^2+b^2}}{1-e}\sin k\tau, \frac{c}{1-e}, \frac{d}{1-e}\right).\end{aligned}$$

Define $G:\left(S^4-\{(0,0,0,0,1)\}\right)\times I\to\mathbb{H}$ by

$$G(a,b,c,d,e,t) = \left(\frac{\sqrt{a^2+b^2}}{1-e}\cos k\tau, \frac{\sqrt{a^2+b^2}}{1-e}\sin k\tau, \frac{c(1-t)}{1-e}, \frac{d(1-t)}{1-e}\right).$$

where τ is determined by

$$\cos\tau = \frac{a}{\sqrt{a^2+b^2}} \text{ and } \sin\tau = \frac{b}{\sqrt{a^2+b^2}}$$

if $a^2+b^2>0$ and $\tau=0$ otherwise.

Then *G* is continuous since its component functions are continuous.

When $t=0$,

$$\begin{aligned}G\left((a,b,c,d,e,0)\right) &= \left(\frac{\sqrt{a^2+b^2}}{1-e}\cos k\tau, \frac{\sqrt{a^2+b^2}}{1-e}\sin k\tau, \frac{c}{1-e}, \frac{d}{1-e}\right).\\ &= \left(p\circ\Sigma^3 s\right)\left((a,b,c,d,e)\right)\end{aligned}$$

When $t=1$,

$$G\left((a,b,c,d,e,1)\right) = \left(\frac{\sqrt{a^2+b^2}}{1-e}\cos k\tau, \frac{\sqrt{a^2+b^2}}{1-e}\sin k\tau, 0, 0\right).$$

Define $v: S^4-\{(0,0,0,0,1)\}\to\mathbb{R}^4$ by

$$v(a,b,c,d,e)=\left(\frac{\sqrt{a^2+b^2}}{1-e}\cos k\tau,\frac{\sqrt{a^2+b^2}}{1-e}\sin k\tau,0,0\right)$$

where τ is determined by $\cos\tau=\dfrac{a}{\sqrt{a^2+b^2}}$ and $\sin\tau=\dfrac{b}{\sqrt{a^2+b^2}}$. Then $\left(p\circ\Sigma^3 s\right)\sim v$.

We now define $H:\left(S^4-\{(0,0,0,0,1)\}\right)\times I\to\mathbb{R}^4$ by

$$\begin{aligned}
&H(a,b,c,d,e,t)\\
&=t\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)\left(\frac{a}{\sqrt{a^2+b^2}},\frac{b}{\sqrt{a^2+b^2}},0,0\right)^k\\
&\quad+(1-t)\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)^k\left(\frac{a}{\sqrt{a^2+b^2+c^2+d^2}},\frac{\sqrt{b^2+c^2+d^2}}{\sqrt{a^2+b^2+c^2+d^2}},0,0\right)^k.
\end{aligned}$$

Then H is continuous since the component functions are continuous as products and sums of continuous functions. Further when $t=0$,

$$\begin{aligned}
&H(a,b,c,d,e,0)\\
&=\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)^k\left(\frac{a}{\sqrt{a^2+b^2+c^2+d^2}},\frac{\sqrt{b^2+c^2+d^2}}{\sqrt{a^2+b^2+c^2+d^2}},0,0\right)^k\\
&=\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)^k(\cos\theta,\sin\theta,0,0)^k\\
&=\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)^k(\cos k\theta,\sin k\theta,0,0)\\
&=u(a,b,c,d,e).
\end{aligned}$$

When $t=1$,

$$\begin{aligned}
H(a,b,c,d,e,1)&=\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)\left(\frac{a}{\sqrt{a^2+b^2}},\frac{b}{\sqrt{a^2+b^2}},0,0\right)^k\\
&=\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)(\cos\tau,\sin\tau,0,0)^k\\
&=\left(\frac{\sqrt{a^2+b^2}}{1-e}\right)(\cos k\tau,\sin k\tau,0,0)=v(a,b,c,d,e).
\end{aligned}$$

Thus $u \sim v$. Since homotopy is an equivalence relation, the fact that

$$\left(x^k \circ p\right) \sim u \sim v \sim \left(p \circ \Sigma^3 s\right)$$

implies

$$\left(x^k \circ p\right) \sim \left(p \circ \Sigma^3 s\right)$$

as desired. ■

Now we have enough to prove Theorem 19.

Proof of Theorem 19: Let $f : \mathbb{H} \to \mathbb{H}$ be defined by $f\left(q\right) = q^k$. Let $p : S^4 \to \mathbb{H} \cup \{\infty\}$ be the stereographic projection of the 5-sphere into the quaternions and let $p^{-1} : \mathbb{H} \cup \{\infty\} \to S^4$ be the one-point compactification of the quaternions into the 5-sphere.

Let $\Sigma^3 s : S^4 \to S^4$ be the 3rd iterated suspension of $s : S^1 \to S^1$ defined by $s\left(z\right) = z^k$. By Theorem 20, $\left(f \circ p\right) \sim \left(p \circ \Sigma^3 s\right)$. Since p^{-1} is a homeomorphism, by Theorem 10

$$\left(p^{-1} \circ f \circ p\right) \sim \left(p^{-1} \circ p \circ \Sigma^3 s\right) = \Sigma^3 s .$$

By Theorem 11, we know homotopic maps induce the same homomorphism on homology groups, so

$$\left(p^{-1} \circ f \circ p\right)_*^n = \left(\Sigma^3 s\right)_*^n .$$

Thus $\deg\left(p^{-1} \circ f \circ p\right) = \deg\left(\Sigma^3 s\right)$. By Theorem 17 and Theorem 18,

$$\deg\left(\Sigma^3 s\right) = \deg\left(s\right) = k .$$

Therefore $\deg\left(p^{-1} \circ f \circ p\right) = k$. ■

We return now to the notion of degree apart from suspension maps.

Theorem 21: If $f : S^n \to S^n$ is not surjective, then $\deg\left(f\right) = 0$.

Proof: Let $f : S^n \to S^n$ be continuous and suppose *f* is not surjective. Then we may choose a point $x_0 \in S^n - f\left(S^n\right)$ and reconsider *f* as a composition $h \circ g$ for functions

$g: S^n \to S^n - \{x_0\}$ and $h: S^n - \{x_0\} \to S^n$. Then by Theorem 9, $f_*^n = h_*^n \circ g_*^n$. As a mapping diagram we have

$$H_n\left(S^n\right) \xrightarrow{g_*^n} H_n\left(S^n - \{x_0\}\right) \xrightarrow{h_*^n} H_n\left(S^n\right).$$

Since $S^n - \{x_0\} \cong \mathbb{R}^n$ it follows $H_n\left(S^n - \{x_0\}\right) = 0$ and therefore, we have $\mathbb{Z} \xrightarrow{g_*^n} 0 \xrightarrow{h_*^n} \mathbb{Z}$. Therefore, both $g_*^n = 0$ and $h_*^n = 0$. Hence, $f_*^n = 0$ and so $\deg(f) = 0$. ■

We use the contrapositive of this statement in the proof of the Fundamental Theorem and so we give it here explicitly. If $\deg(f) \neq 0$ then *f* is surjective.

We may now explicitly turn to polynomials over the quaternions. We begin by defining by induction the algebraic degree of a monomial (to distinguish this notion of degree from the topological notion). A constant function $f(q) = a$ for a fixed $a \in \mathbb{H}$ is defined to have algebraic degree 0. The variable *q* is said to be of algebraic degree 1. Therefore, polynomials of the form $aq, qb,$ or aqb for $a, b \in \mathbb{H}$ are monomials of degree 1. In general, the monomial of the form $a_1 q^{n_1} a_2 q^{n_2} \cdots a_m q^{n_m} a_{m+1}$ with $a_i \in \mathbb{H}$ has algebraic degree $n = n_1 + n_2 + \cdots + n_m$. We define a polynomial over the quaternions to be a finite sum of such monomials.

Definition 13: Let $f: \mathbb{H} \to \mathbb{H}$ be a polynomial in quaternions. The *algebraic degree of f*, denoted $\operatorname{alg}\deg(f)$, is the maximum of the degrees of the finitely many monomials whose sum is *f*.

Example 8: Let $f(q) = (3 + 4i + 5j)q^2(4j + 6k)q + q(1 + i + j + k)q + (5i - k)$. Then *f* is a polynomial over the quaternions since it is the sum of a finite number of monomials over the quaternions. The degrees of the first, second, and third monomials are 3, 2, and 0, respectively. Therefore $\operatorname{alg}\deg(f) = 3$.

Recall that a polynomial in quaternions can have two monomials of the same degree. For example, consider $g(q) = iq - qi$. Each monomial has degree 1, therefore *g* is a function of degree 1. The Fundamental Theorem for Polynomials in Quaternions requires that a function have a degree that is determined by a *unique* leading monomial.

Definition 14: Let f be a polynomial over the quaternions and suppose $\operatorname{alg}\deg(f) = n$. Then f is the sum of a finite number of monomials at least one of whose degree is *n*. If there is only one monomial in the sum whose degree is *n*, then f is said to be a polynomial over the quaternions with *a unique leading monomial of degree n*.

Theorem 22: Let $f:\mathbb{H}\to\mathbb{H}$ be a polynomial in quaternions with a unique leading monomial of degree *n*. Let $g(q)=q^n$. Then f and *g* are homotopy equivalent.

Proof: The proof is rather straightforward and is presented fully in [3]. In fact, the same technique that is used to prove the Fundamental Theorem of Algebra in $\mathbb{C}$ is used here with little modification.

Define $f(q)=a_0qa_1qa_2q\cdots qa_{n-1}qa_n+\varphi(q)$ where $a_0,a_1,\ldots,a_n\in\mathbb{H}-\{0\}$ and $\varphi(q)$ is a finite sum of monomials $b_0qb_1q\cdots qb_k$ with $k<n$. It is clear that $\varphi(q)$ is continuous since it is the sum of continuous monomials. Define $F:\mathbb{H}\times I\to\mathbb{H}$ by

$$F(q,t)=a_0qa_1qa_2q\cdots qa_{n-1}qa_n+(1-t)\varphi(q).$$

Then *F* is continuous in *q* since, with *t* fixed, *F* is the sum of monomials. Similarly, *F* is continuous in *t* since, with *q* fixed, *F* is simply a linear function in $\mathbb{H}$. Hence *F* is continuous. Further $F(q,0)=f(q)$ and $F(q,1)=a_0qa_1qa_2q\cdots qa_{n-1}qa_n$. Let $\gamma(q)=a_0qa_1qa_2q\cdots qa_{n-1}qa_n$. Then f is homotopic to γ.

Now it remains to show that γ is homotopic to $g(q)=q^n$. To this end, we note that $\mathbb{H}-\{0\}$ is path connected. For $0\le i\le n$, let $\alpha_i(t)$ be a path from a_i to 1 in $\mathbb{H}-\{0\}$ where $0\le t\le 1$ such that $\alpha_i(0)=a_i$ and $\alpha_i(1)=1$. Define

$$\beta(q,t)=\alpha_0(t)q\alpha_1(t)q\alpha_2(t)q\cdots q\alpha_{n-1}(t)q\alpha_n(t).$$

Then β is continuous in both *t* and *q*. Further $\beta(q,0)=\gamma(q)$ and $\beta(q,1)=q^n=g(q)$. Thus β is a homotopy between γ and *g*. Finally, since homotopy equivalence is an equivalence relation, we conclude that $f\sim g$. ■

(Theorem 23) The Fundamental Theorem of Algebra for Polynomials in Quaternions: Let $f:\mathbb{H}\to\mathbb{H}$ be a continuous polynomial in quaternions with a unique leading monomial of nonzero degree *d*. That is, $\operatorname{alg}\deg(f)=d>0$. Then f has at least one zero in $\mathbb{H}$.

Proof: Let $f:\mathbb{H}\to\mathbb{H}$ be a continuous monic polynomial with $\operatorname{alg}\deg(f)=d>0$. Let $\left(p^{-1}\circ f\circ p\right):S^4\to S^4$ where p^{-1} and p are, respectively, the one-point compactification of $\mathbb{H}\cup\{\infty\}$ to S^4 and the stereographic projection of S^4 to $\mathbb{H}\cup\{\infty\}$.

By Theorem 22, $f\sim x^d$. By Theorem 10, $\left(p^{-1}\circ f\circ p\right)\sim\left(p^{-1}\circ x^d\circ p\right)$. By Theorem 11 $\deg\left(p^{-1}\circ f\circ p\right)=\deg\left(p^{-1}\circ x^d\circ p\right)$. And by Theorem 19, $\deg\left(p^{-1}\circ x^d\circ p\right)=d$. Hence $\deg\left(p^{-1}\circ f\circ p\right)=d$. Since $d\neq 0$, by the contrapositive of Theorem 21 the map $p^{-1}\circ f\circ p$ is surjective. So, there exists some $x\in S^4$ such that $p^{-1}\left(f\left(p(x)\right)\right)=0$. From this we derive $p(0)=p\left(p^{-1}\left(f\left(p(x)\right)\right)\right)=f\left(p(x)\right)$. The stereographic projection sends 0 to 0, whence, $0=f\left(p(x)\right)$. Hence, there exists some $q\in\mathbb{H}$ such that $0=f(q)$. That is, f has a root in $\mathbb{H}$. ■

We note that there is nothing particularly special about the quaternions in this case. A slight modification of Theorem 23 would apply in the case where $f:\mathbb{R}^n\to\mathbb{R}^n$ is a polynomial in $\mathbb{R}^n$ such that $\lim_{|x|\to\infty}\left|f(x)\right|=\infty$. To prove this stronger case we would simply extend the definition of unique leading monomials of degree *n* to polynomials over $\mathbb{R}^n$ and then rephrase some of our theorems to fit the context of $\mathbb{R}^n$ rather than $\mathbb{H}$.

The Fundamental Theorem of Algebra for Polynomials in Quaternions was proved for the first time in 1941 by Eilenberg and Niven. To some, it is quite surprising that this result appears for the first time roughly a century after the discovery of quaternions. To others, it is not surprising at all since the proof requires the techniques of algebraic topology developed much later. In fact, as far as we know ([6]), this result has not been proven with techniques outside of topology.

3. Frobenius's Theorem on Associative Division Algebras

Definition 15: A vector space *V* over $\mathbb{R}$ with a product map $V \times V \to V$ by $(x, y) \mapsto xy$ (i.e. there is some rule indicating how to multiply two elements in *V*) is *an algebra over* $\mathbb{R}$ if it has the left and right distributive property. That is, for all $\alpha, \beta \in \mathbb{R}$ and all $x, y, z \in V$,

$$x(\alpha y + \beta z) = x(\alpha y) + x(\beta z) = \alpha(xy) + \beta(xz)$$

and

$$(\alpha x + \beta y)z = (\alpha x)z + (\beta yz) = \alpha(xz) + \beta(yz).$$

Definition 16: A non-zero algebra *A* is called a *division algebra* if *A* has an identity element and for all $a, b \in A$ with *a* nonzero, the equations $ax = b$ and $ya = b$ have unique solutions in *A*.

Note: We will assume that all division algebras under consideration below are finite-dimensional and associative. That is, if *A* is a division algebra and *B* is a basis for *A*, then $\dim(B) < \infty$, and for all $x, y, z \in A$, $x(yz) = (xy)z$.

Theorem 24: $\mathbb{H}$ is a division algebra.

Proof: It is rather trivial that $\mathbb{H}$ is a vector space over $\mathbb{R}$. After all, for $x, y \in \mathbb{H}$ and $\alpha, \beta \in \mathbb{R}$ we already know that $\alpha x + \beta y \in \mathbb{H}$. Furthermore, there exist additive and multiplicative identities, 0 and 1 respectively.

Let $x, y, z \in \mathbb{H}$ and $\alpha, \beta \in \mathbb{R}$. Using the fact that the function Φ in Theorem 1 is a ring homomorphism, we use the distributive property of 2×2 matrices with complex entries to obtain

$$\begin{aligned}\Phi(x(\alpha y + \beta z)) &= \Phi(x)\Phi(\alpha y + \beta z) \\ &= \Phi(x)(\Phi(\alpha y) + \Phi(\beta z)) \\ &= \Phi(x)\Phi(\alpha y) + \Phi(x)\Phi(\beta y).\end{aligned}$$

Recalling that real numbers commute with all quaternions and that multiplication is associative on the quaternions, we derive

$$\Phi(x)\Phi(\alpha y) + \Phi(x)\Phi(\beta y) = \Phi(x(\alpha y)) + \Phi(x(\beta y)) = \Phi(\alpha(xy) + \beta(xy)).$$

Since Φ is a bijection, it follows that $x(\alpha y+\beta z)=\alpha(xy)+\beta(xz)$. Hence, the left distributive property holds. A nearly identical argument establishes that the right distributive property also holds. Thus $\mathbb{H}$ is a division algebra. ■

Definition 17: Let *A* be an algebra with identity element *e*. The *imaginary space of A*, denoted by $\operatorname{im}A$, is the set $\operatorname{im}A=\left\{x\in A \middle| x^2\in\mathbb{R}e \text{ and } x\notin\mathbb{R}e-\{0\}\right\}$.

Theorem 25: Let *A* be a nonzero associative division algebra with identity element *e*. Suppose also that $U\subseteq\operatorname{im}A$ is a two-dimensional vector subspace of *A*. Then for every $p\in U$, there exists some $q\in U-\mathbb{R}p$ such that $pq+qp=0$.

Proof: Let *A* be a nonzero associative division algebra such that U is a two-dimensional vector subspace of *A* that is contained in $\operatorname{im}A$ with basis $\{u,v\}$. Since *U* is a vector space, $u+v\in U$. Therefore, $(u+v)^2\in\mathbb{R}e,\ u^2\in\mathbb{R}e,$ and $v^2\in\mathbb{R}e$. So

$$uv+vu=(u+v)^2-u^2-v^2\in\mathbb{R}e.$$

Let $p\in U$ with $p\neq 0$, and let $\lambda\in\mathbb{R}-\{0\}$ such that $p^2=\lambda e$. There exist scalars $\alpha,\beta\in\mathbb{R}$ such that $p=\alpha u+\beta v$. Since $\dim(U)=2$, there exists $x\in U$ such that *p* and *x* are linearly independent. Then, $x=\gamma u+\delta v$ for some scalars $\gamma,\delta\in\mathbb{R}$. Next we consider

$$\begin{aligned}px+xp&=(\alpha u+\beta v)(\gamma u+\delta v)+(\gamma u+\delta v)(\alpha u+\beta v)\\&=\alpha\gamma u^2+\alpha\delta uv+\beta\gamma vu+\beta\delta v^2+\gamma\alpha u^2+\gamma\beta uv+\delta\alpha vu+\delta\beta v^2\\&=2\alpha\gamma u^2+2\beta\delta v^2+(\alpha\delta+\beta\gamma)(uv+vu).\end{aligned}$$

Hence we observe that $px+xp\in\mathbb{R}e$ since $u^2,\ v^2$ and $uv+vu\in\mathbb{R}e$. So $px+xp=\theta e$ for some $\theta\in\mathbb{R}$. Define $q=x-\frac{\theta}{2\lambda}p$. Then, $q\in U-\mathbb{R}p$ since *x* and *p* are linearly independent, and

$$\begin{aligned}pq+qp&=p\left(x-\frac{\theta}{2\lambda}p\right)+\left(x-\frac{\theta}{2\lambda}p\right)p=px-\frac{\theta}{2\lambda}p^2+xp-\frac{\theta}{2\lambda}p^2\\&=px+xp-\frac{\theta}{\lambda}p^2=px+xp-\frac{\theta\lambda e}{\lambda}=px+xp-\theta e=0\end{aligned}$$

as desired. ■

Definition 18: Let $f(x)=\sum_{i=0}^{n} r_i x^i \in \mathbb{R}[x]$ and let *A* be a nontrivial division algebra. For $a \in A$, we define $f(a)=\sum_{i=0}^{n} r_i a^i$ and we note $f(a) \in A$. We refer to this as the *substitution rule.*

Definition 19: Let *A* be a nontrivial division algebra. For a fixed $a \in A$ we define *the substitution function of a,* $\varphi_a : \mathbb{R}[x] \to A$ by $\varphi_a(f(x)) = f(a)$. Note that this function is well-defined since *A* is an associative algebra.

Theorem 26: Let *A* be a nontrivial division algebra. For each $a \in A$ the substitution function of *a* is a homomorphism.

Proof: Fix $a \in A$ and let $f(x), g(x) \in \mathbb{R}[x]$. It follows from the substitution rule that

$$\varphi_a(f(x)+g(x)) = f(a)+g(a) = \varphi_a(f(x)) + \varphi_a(g(x))$$

and

$$\varphi_a(f(x)g(x)) = f(a)g(a) = \varphi_a(f(x))\varphi_a(g(x)).$$

And so φ_a is a homomorphism for each $a \in A$.

Theorem 27: Let *A* be a nontrivial associative division algebra. Then, for every $a \in A$ there exist scalars $\theta, \lambda \in \mathbb{R}$ such that $a^2 = \lambda + \theta a$.

Proof: Let *A* be a nontrivial associative division algebra and fix $a \in A$. Because φ_a is a homomorphism, $\ker \varphi_a$ is an ideal of $\mathbb{R}[x]$. In particular, $\ker \varphi_a$ is not the trivial ideal since $\dim A < \infty$.

Let $p(x)$ be the unique monic polynomial of lowest degree in $\mathbb{R}[x]$ such that $p(a)=0$. (Such a nonzero polynomial exists since the ideal under consideration is nontrivial). Since complex roots of polynomials with real coefficients come in conjugate pairs, there must be an even number of complex roots. If the degree of *p* is odd, there must therefore be at least one real root. So $p(x)=(x-\mu)q(x)$ for some $\mu \in \mathbb{R}$ and $q(x) \in \mathbb{R}[x]$ with $\deg(q(x)) < \deg(p(x))$. Since *p* is the unique monic polynomial of lowest degree such that $p(a)=0$, it follows that $q(a) \neq 0$. Therefore $a-\mu=0$. Hence $a^2 = \lambda + \theta a$ for scalars $\theta = 0$ and $\lambda = \mu^2$.

If $\deg(p(x))$ is even, then there exist a pair of conjugate complex roots ξ and $\overline{\xi}$ of *p*. Then

$$p(x) = (x-\xi)(x-\overline{\xi})r(x) = \left(x^2 - \left(\xi+\overline{\xi}\right)x + \xi\overline{\xi}\right)r(x)$$

where $r(x) \in \mathbb{R}[x]$ with $\deg(r(x)) < \deg(p(x))$. Again since *p* is the unique monic polynomial of lowest degree with $p(a)=0$, $r(a) \neq 0$. Hence $a^2 - \left(\xi+\overline{\xi}\right)a + \xi\overline{\xi} = 0$. It is clear that $\xi+\overline{\xi}$ and $\xi\overline{\xi}$ are both real numbers. Setting $\theta = \xi+\overline{\xi}$ and $\lambda = -\xi\overline{\xi}$, we have $a^2 = \lambda + \theta a$.

Since these are the only possible cases and since *a* was arbitrary, it follows that for all $a \in A$ there exist scalars $\theta, \lambda \in \mathbb{R}$ so that $a^2 = \lambda + \theta a$. ■

Theorem 28: For a nonzero associative division algebra *A*, $\operatorname{im} A$ is a vector subspace of *A* and $A = \mathbb{R}e \oplus \operatorname{im} A$.

Proof: Let *A* be a nonzero associative division algebra. First, note that $\operatorname{im} A \neq \varnothing$ since $0 \in \operatorname{im} A$. Also, if $u \in \operatorname{im} A$, then $u^2 \in \mathbb{R}e$.

So for any $\alpha \in \mathbb{R}$, $(\alpha u)^2 = \alpha u \alpha u = \alpha^2 u^2 \in \mathbb{R}e$. Therefore, $\alpha u \in \operatorname{im} A$, and so $\operatorname{im} A$ is closed under scalar multiplication.

Let $u, v \in \operatorname{im} A$. If *u* and *v* are linearly dependent then either *u* or *v* is a scalar multiple of the other and so $\theta u + \lambda v \in \operatorname{im} A$ for all scalars $\theta, \lambda \in \mathbb{R}$, since $\operatorname{im} A$ is closed under scalar multiplication.

Suppose *u* and *v* are linearly independent. Suppose $u + v = r$ for some $r \in \mathbb{R}e - \{0\}$ (since *u* and *v* are linearly independent it cannot be true that $u + v = 0$). Then it follows that

$$u^2 = (r-v)^2 = r^2 - rv - vr + v^2.$$

Therefore, $u^2 - r^2 - v^2 = -2rv$. However, the left hand side of this equation is in $\mathbb{R}e$, but the right hand side is an element of $\operatorname{im} A$. Since $\mathbb{R}e \cap \operatorname{im} A = \{0\}$ it follows that both $u^2 - r^2 - v^2$ and $-2rv$ must be 0. However since $v \neq 0$, $-2rv \neq 0$ and thus we have a contradiction.

Hence, $u + v \notin \mathbb{R}e$. Therefore, in order to show that $\operatorname{im} A$ is a vector subspace of *A* it suffices to show that $(u+v)^2 \in \mathbb{R}e$.

By Theorem 27, there exist scalars $\theta,\lambda,\sigma,\tau\in\mathbb{R}$ such that $(u+v)^2=\theta+\lambda(u+v)$ and $(u-v)^2=\sigma+\tau(u-v)$.

Thus,

$$\begin{aligned}u^2+uv+vu+v^2+u^2-uv-vu+v^2 &= (u+v)^2+(u-v)^2\\ &= \theta+\lambda(u+v)+\sigma+\tau(u-v)\\ &= \theta+\sigma+(\lambda+\tau)u+(\lambda-\tau)v,\end{aligned}$$

whence we have

$$2\left(u^2+v^2\right)-(\theta+\sigma)=(\lambda+\tau)u+(\lambda-\tau)v$$

to which we will refer as $(*)$.

Since the left hand side of $(*)$ is a real number, so too is the right hand side of $(*)$. Let $\gamma=(\lambda+\tau)u+(\lambda-\tau)v$. Then $\gamma\in\mathbb{R}e$. Since $u^2,v^2\in\mathbb{R}e$, there are scalars α and β such that $u^2=\alpha e$ and $v^2=\beta e$. Then

$$\begin{aligned}\gamma^2 &= \left((\lambda+\tau)u+(\lambda-\tau)v\right)^2\\ &= (\lambda+\tau)^2u^2+(\lambda+\tau)(\lambda-\tau)uv+(\lambda+\tau)(\lambda-\tau)vu+(\lambda-\tau)^2v^2\\ &= (\lambda+\tau)^2\alpha e+\left(\lambda^2-\tau^2\right)(uv+vu)+(\lambda-\tau)^2\beta e\\ &= \left((\lambda+\tau)^2\alpha+(\lambda-\tau)^2\beta\right)e+\left(\lambda^2-\tau^2\right)(uv+vu).\end{aligned}$$

We now have two cases.

Case 1: If $\lambda^2-\tau^2=0$, then either $\lambda-\tau=0$ or $\lambda+\tau=0$.

> Case 1(a): If $\lambda-\tau=0$, we would deduce from equation $(*)$ that $(\lambda+\tau)u\in\mathbb{R}e$, which implies $\lambda+\tau=0$, since *u* and *v* are linearly independent and so $u\neq 0$.
>
> Case 1(b): If $\lambda+\tau=0$, we would deduce from equation $(*)$ that $(\lambda-\tau)v\in\mathbb{R}e$, which implies $\lambda-\tau=0$.

Therefore we conclude that $\lambda^2-\tau^2=0$ if and only if $\lambda=\tau=0$. In this case, $(u+v)^2=\theta\in\mathbb{R}e$.

Case 2: If $\lambda^2-\tau^2\neq 0$ then

$$uv+vu=\left(\lambda^2-\tau^2\right)^{-1}\left[\gamma^2-\left(\left(\lambda+\tau\right)^2\alpha+\left(\lambda-\tau\right)^2\beta\right)e\right]\in\mathbb{R}e\,.$$

Since $uv+vu,\ u^2,$ and v^2 are elements of $\mathbb{R}e$, the sum $uv+vu+u^2+v^2\in\mathbb{R}e$. But note that $\left(u+v\right)^2=u^2+uv+vu+v^2$. Thus $\left(u+v\right)^2\in\mathbb{R}e$.

Hence, in both cases we can conclude that $u+v\in\operatorname{im}A$. Therefore $\operatorname{im}A$ is indeed a vector subspace of A.

Now suppose $x\in A-\mathbb{R}e$. By Theorem 27 there exist scalars $\theta,\lambda\in\mathbb{R}$ such that $x^2=\lambda+\theta x$. Observe that $\left(x-\dfrac{\theta}{2}\right)^2=x^2-\theta x+\dfrac{\theta^2}{4}=\lambda+\dfrac{\theta^2}{4}\in\mathbb{R}e$. Since $x\notin\mathbb{R}e$, $x-\dfrac{\theta}{2}\notin\mathbb{R}e$. Thus, $x-\dfrac{\theta}{2}\in\operatorname{im}A$. Let $u=x-\dfrac{\theta}{2}$. Then $x=\dfrac{\theta}{2}+u$ with $\dfrac{\theta}{2}\in\mathbb{R}$ and $u\in\operatorname{im}A$. Hence $x\in\mathbb{R}e+\operatorname{im}A$. Since x was arbitrary, $A-\mathbb{R}e\subseteq\mathbb{R}e+\operatorname{im}A$. Further, it is clear that $\mathbb{R}e\subseteq\mathbb{R}e+\operatorname{im}A$.

Therefore $A=\mathbb{R}e\cup\left(A-\mathbb{R}e\right)\subseteq\mathbb{R}e+\operatorname{im}A$. Since $\mathbb{R}e+\operatorname{im}A\subseteq A$, $A=\mathbb{R}e+\operatorname{im}A$. Finally, $\mathbb{R}\cap\operatorname{im}A=\{0\}$. Thus $A=\mathbb{R}\oplus\operatorname{im}A$. ■

Definition 20: Let A be an algebra with identity element. We call three elements $u,v,w\in A$ a *Hamiltonian triple* if u, v, and w satisfy the equations

$$u^2=v^2=w^2=-1,\quad w=uv=-vu,\quad u=vw=-wv,\ \text{and}\quad v=wu=-uw,$$

where we have used -1 to denote the additive inverse of the identity.

Theorem 29: Let A be a nonzero associative division algebra, and let U be a two-dimensional vector subspace of the imaginary space of A. Then for every $u\in U$ with $u^2=-1$, there exists a $v\in U$ such that $u,v,$ and uv form a Hamiltonian triple. Further u, v, and $w=uv$ are linearly independent.

Proof: Let A be a nonzero associative division algebra, and let U be a two-dimensional vector subspace of the imaginary space of A. Let $u\in U$ with $u^2=-1$. From Theorem 25, there exists some nonzero $s\in U$ such that $us+su=0$. Since $s\in U\subseteq\operatorname{im}A$, $s\notin\mathbb{R}$, and $s^2=\alpha$ for some $\alpha\in\mathbb{R}$ with $\alpha<0$.

Set $v=\dfrac{s}{\sqrt{-\alpha}}$ to obtain

$$uv + vu = u\left(\frac{s}{\sqrt{-\alpha}}\right) + \left(\frac{s}{\sqrt{-\alpha}}\right)u = \frac{1}{\sqrt{-\alpha}}\left(us + su\right) = 0$$

and

$$v^2 = \frac{s^2}{-\alpha} = -1.$$

Now let $w = uv = -vu$. Then

$$w^2 = \left(uv\right)^2 = u\left(vu\right)v = u\left(-uv\right)v = -u^2v^2 = -1.$$

Also

$$vw = v\left(-vu\right) = -v^2u = u$$

and

$$wv = \left(uv\right)v = uv^2 = -u ,$$

Whence $u = vw = -wv$. Similarly,

$$wu = \left(-vu\right)u = -vu^2 = v$$

and

$$uw = u\left(uv\right) = u^2v = -v ,$$

whence $v = wu = -uw$. Thus, *u*, *v*, and *w* are, indeed, a Hamiltonian triple.

Finally, to prove that *u*, *v*, and *w* are linearly independent, assume $0 = \alpha u + \beta v + \gamma w$ for scalars $\alpha, \beta, \gamma \in \mathbb{R}$. Then we have the following implications:

$$\begin{aligned} & 0 = \alpha u + \beta v + \gamma w \\ \Rightarrow\ & 0 = 0w = \left(\alpha u + \beta v + \gamma w\right)w = \alpha uw + \beta vw + \gamma w^2 = -\alpha v + \beta u - \gamma \\ \Rightarrow\ & \gamma = -\alpha v + \beta u, \end{aligned}$$

whence

$$\begin{aligned} \gamma^2 &= \left(-\alpha v + \beta u\right)^2 \\ &= \alpha^2 v^2 - \alpha\beta vu - \alpha\beta uv + \beta^2 u^2 \\ &= -\alpha^2 - \alpha\beta\left(vu + uv\right) - \beta^2 \\ &= -\left(\alpha^2 + \beta^2\right). \end{aligned}$$

Thus $\alpha = \beta = \gamma = 0$. Hence *u*, *v*, and *w* are linearly independent. ■

Theorem 30: Let *A* be a nonzero associative division algebra with $\dim A \geq 3$. Then *A* contains a Hamiltonian triple and therefore a subalgebra *B* such that $B \cong \mathbb{H}$.

Proof: Let *A* be a nonzero associative division algebra with $\dim A \geq 3$. By Theorem 28, $A = \mathbb{R} \oplus \operatorname{im} A$. Therefore, $\dim(\operatorname{im} A) = \dim A - \dim \mathbb{R} \geq 3 - 1 = 2$. Since $\operatorname{im} A$ is a vector subspace of *A* of dimension at least two, there exists a two-dimensional vector subspace $U \subseteq \operatorname{im} A$. Suppose $y \in U$ such that $y^2 = \alpha \in \mathbb{R}$. Again since $y \in \operatorname{im} A$, $\alpha < 0$. Since *U* is a vector subspace of $\operatorname{im} A$, it is closed under scalar multiplication.

Therefore $\frac{y}{\sqrt{-\alpha}} \in U$. Finally, $\left(\frac{y}{\sqrt{-\alpha}}\right)^2 = \frac{y^2}{-\alpha} = -1$. Therefore we may apply Theorem 29 to obtain a Hamiltonian triple $\{u, v, w\}$ in *U* and hence in *A*.

Define $B = \mathbb{R} \oplus \mathbb{R}u \oplus \mathbb{R}v \oplus \mathbb{R}w$ and define a map $f : B \to \mathbb{H}$ by

$$f(\alpha + \beta u + \gamma v + \delta w) = \alpha + \beta i + \gamma j + \delta k.$$

Then, since *u*, *v*, and *w* are linearly independent, f is easily seen to be injective and surjective. The fact that *f* is a homomorphism follows from the fact that *u*, *v*, and *w* satisfy the nine Hamiltonian relations that *i*, *j*, and *k* satisfy.

Hence, f is an isomorphism. ■

Theorem 31 (Frobenius's Theorem): Let *A* be a nonzero associative division algebra. Then *A* is isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$.

Proof: Let *A* be a nonzero associative division algebra. We divide into cases.

Case 1: Suppose $\dim A = 1$. Then the multiplicative identity may serve as a basis for *A*. Denote this identity by *e*. Define $f : \mathbb{R} \to A$ by $f(r) = re$. Then f is bijective since each element of *A* can be expressed *uniquely* as a scalar multiple of *e*. Also, *f* is a ring homomorphism since for $r, s \in \mathbb{R}$,

$$f(r+s) = (r+s)e = re + se = f(r) + f(s).$$

And,

$$f(r)f(s) = rese = rse^2 = rse = f(rs).$$

Thus *f* is an isomorphism and $\mathbb{R} \cong A$.

Case 2: Now suppose $\dim A = 2$. Since $A = \mathbb{R} \oplus \operatorname{im} A$, then $\dim(\operatorname{im} A) = 1$. By our discussion in the proof of Theorem 29, we know there exists $u \in \operatorname{im} A$ such that $u^2 = -1$. We then define $f : \mathbb{C} \to A$ by $x + yi \mapsto x + yu$. We see trivially that f is injective and surjective. Since u satisfies the same relationship as i, f is a homomorphism. Thus, $\mathbb{C} \cong A$.

Case 3: Now suppose $\dim A \geq 3$. Then by Theorem 29, A contains a Hamiltonian triple $u, v, w \in \operatorname{Im} A$ with $w = uv$. Let $x \in \operatorname{im} A$. Observe that $ux + xu = (u + x)^2 - u^2 - x^2$. Since $x, u \in \operatorname{im} A$ and $\operatorname{im} A$ is a vector space, $u + x \in \operatorname{im} A$. So the right hand side of the above equation is a real number. Therefore, $xu + ux = \alpha$ for some $\alpha \in \mathbb{R}$. Similarly there are $\beta, \gamma \in \mathbb{R}$ such that $xv + vx = \beta$ and $xw + wx = \gamma$. Thus,

$$\alpha v = (xu + ux)v = xuv + uxv = xw + uxv$$

and

$$u\beta = u(xv + vx) = uxv + uvx = uxv + wx.$$

Thus,

$$\alpha v - \beta u = xw + uxv - uxv - wx = xw - wx.$$

And then,

$$\alpha v - \beta u + \gamma = xw - wx + xw + wx = 2xw,$$

so that

$$(\alpha v - \beta u + \gamma)w = 2xww = -2x.$$

Thus,

$$x = \frac{-\alpha}{2}u + \frac{-\beta}{2}v + \frac{-\gamma}{2}w.$$

And so $x \in \mathbb{R}u \oplus \mathbb{R}v \oplus \mathbb{R}w$. Thus we have shown that $\operatorname{im} A \subseteq \mathbb{R}u \oplus \mathbb{R}v \oplus \mathbb{R}w$. Therefore, $A \subseteq \mathbb{R} \oplus \mathbb{R}u \oplus \mathbb{R}v \oplus \mathbb{R}w$. The reverse inclusion is trivial and so by Theorem 30 $A \cong \mathbb{R} \oplus \mathbb{R}u \oplus \mathbb{R}v \oplus \mathbb{R}w \cong \mathbb{H}$. ■

Conclusion

With the proof of Frobenius's Theorem, we end our exploration of the quaternion algebra. Frobenius's Theorem serves as a pleasant capstone since its implication is that our study of quaternions throughout this paper cannot be extended into higher dimensions without sacrificing at least one property. It is known that if the requirement of multiplicative associativity is dropped, then we can find one more division algebra of dimension 8, namely the octonians ([4]). Of course, without associativity care must be exercised when performing calculations. Our proof of the Fundamental Theorem of Algebra for Quaternions extends rather easily to the octonians since we did not make use of associativity in that proof. However, note that polynomials in octonians are thoroughly frustrating due to the lack of associativity of coefficients ([3]). For example, let a and b denote octonians. Then, although the polynomials $(ax)b$ and $a(xb)$ have the same octonian coefficients, they are entirely different. It is also worth pointing out that Frobenius's Theorem prevents us from finding an associative division algebra of dimension 8 by modifying Theorem 1**.** That is, it is impossible to construct an isomorphism from the set of two by two matrices with quaternion entries into the octonians since matrix multiplication in the former set is associative while multiplication in the latter set is not. Thus, Frobenius's Theorem allows us to leave quaternion analysis with a sense of completion or at least contentment that we have done most of what we can with the quaternion algebra.

Works Cited

[1] Boyer, Carl B., and Uta C. Merzbach. *A History of Mathematics*. Hoboken, NJ, Wiley, 2011.

[2] Ebbinghaus, Heinz-Dieter, and John H. Ewing. *Numbers*. New York, Springer, 1995.

[3] Eilenberg, and Niven, Ivan. "The Fundamental Theorem of Algebra for Quaternions." *Bulletin of the American Mathematical Society*, vol. 50, 1944, pp. 246 – 248.

[4] Hatcher, Allen. *Algebraic Topology*. Cambridge, Cambridge University Press, 2015.

[5] Massey, William S. *A Basic Course in Algebraic Topology*. p. 189. Springer, 1991. Available at https://books.google.com/books?id=svbU9nxi2xQC&printsec=frontcover#v=onepage&q&f= false. Accessed on October 22, 2017.

[6] Niven, Ivan. "Extension of a Topological Proof of the Fundamental Theorem of Algebra." *The American Mathematical Monthly*, vol. 57, no. 4, 1950, pp. 246–248., doi:10.2307/2305939.